

Command line parameters:

LOCUS	HSU76705	4155 bp	mrna	PRI	26-JAN-1999
DEFINITION	Human putative RNA binding protein Koc1 mRNA, complete cds.				
ACCESSION	U76705				
VERSION	U76705.1	GI:4098296			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 4155)				
AUTHORS	Mueller-Pillasch, F., Lacher, U. and Wallraab, C.				

JOURNAL
Submitted (30 Oct 1998), Innere Medizin 1, University of Ulm,
Robert-Koch-Str.8, Ulm 89081, Germany

FEATURES
Location/Qualifiers
1. 4155
source

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/chromosome="7"
```

```

/map="7p11.5"
/cell_line="Patu898s"
/tissue_type="pancreas"
251. 1990
CDS

```

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
9b_pr:HSU76705	+	2956.00	3401.17	3.6e-181	4155	! U76705 Human putative RNA bind
9b_pr:HSU97188	+	2956.00	3401.11	3.6e-181	4181	! U97188 Homo sapiens putative R
9b_pr:AB0461108	+	2949.00	3393.95	1.0e-180	4171	! AF117108 Homo sapiens IGF-II m
9b_ro:AB0461108	+	2948.00	3393.94	2.6e-175	1740	! AB0461108 Mus musculus IGF-II m
9b_pr:HS497J21	+	2956.00	3254.68	2.7e-156	129432	! AL0233775 Human DNA sequence
9b_ov:AF064633	+	2483.00	2958.92	5.7e-151	2697	! AF064633 Xenopus laevis Vgl RN
9b_ov:AF042353	+	2481.00	2957.51	6.9e-151	3147	! AF042353 Xenopus laevis KH dom
9b_ov:AF055923	+	2461.50	2836.11	1.1e-149	2161	! AF055923 Xenopus laevis RNA bi
9b_ov:AF064634	+	2461.50	2835.13	1.2e-149	2405	! AF064634 Xenopus laevis RNA bi
9b_ov:AF161270	+	2261.50	2605.82	7.6e-134	2021	! AF161270 Dario rerio Vgl RNA b
9b_ov:AF162527	+	2219.50	2557.77	3.6e-134	2021	! AF026527 Gallus gallus zfpcode
9b_ro:AF061569	+	2192.00	2524.65	2.4e-132	2223	! AF061569 Mus musculus coding r
9b_pr:AF117106	+	2191.00	2523.88	2.6e-132	2230	! AF117106 Homo sapiens IGF-II m
9b_pr:AF117109	+	2181.00	2511.32	1.3e-131	2381	! AF117109 Homo sapiens mRNA-bi
9b_pr:AF1198254	+	1950.50	2246.70	7.4e-117	2010	! AF1198254 Homo sapiens IGF-II m
9b_pr:AF057352	+	1868.00	2145.95	3.0e-111	3667	! AF057352 Homo sapiens hepatoc
9b_pr:AC020629	+	1661.50	1874.09	4.1e-96	142971	! AC020629 Homo sapiens 22q BA
9b_htg:AC015706	+	1556.50	1570.61	3.1e-89	182695	! AC015706 Homo sapiens chromo
9b_htg:AL596372	+	1556.50	1570.51	3.1e-89	184653	! AL596372 Homo sapiens chromo
9b_htg:AL139152	+	1549.50	1748.62	4.0e-89	93682	! AL139152 Homo sapiens chromo
9b_htg:AC024199	+	1286.00	1437.98	8.0e-72	188273	! AC024199 Homo sapiens chromo
9b_htg:AC092447	+	1281.00	1432.55	1.6e-71	181369	! AC092447 Homo sapiens chromo
9b_ro:HUSB1C	+	1040.00	1193.55	3.3e-58	2437	! L35549 Mus musculus Y-box bind
9b_htg:AC069053	+	1013.50	1133.78	7.1e-45	59747	! AC069053 Homo sapiens chromo
9b_in:AF241237	-	835.00	954.38	7.0e-45	3188	! AF241237 Drosophila melanogast
9b_htg:AC012929	+	742.50	823.11	1.4e-37	46580	! AC012929 Drosophila melanogast
9b_in:AE003484	-	742.50	805.71	1.3e-36	314087	! AE003484 Drosophila melanogast
9b_htg:AC090245	+	676.50	745.02	3.2e-33	57261	! AC090245 Homo sapiens chromo
9b_htg:AC026579	+	616.50	674.19	2.8e-29	67923	! AC026579 Homo sapiens chromo
9b_htg:AC026579	+	523.00	566.22	2.9e-23	67923	! AC026579 Homo sapiens chromo
9b_pr:AC021876	+	480.00	512.61	2.8e-20	104668	! AC021876 Homo sapiens clone
9b_htg:AC023375	+	480.00	508.85	4.6e-20	158105	! AC023375 Homo sapiens chromo
9b_ro:AC084407	+	376.50	385.54	3.4e-13	239837	! AC084407 Mus Musculus Strain
9b_pr:AC005082	+	367.50	378.30	8.5e-13	169739	! AC005082 Homo sapiens clone
9b_htg:AC090865	+	305.00	317.07	2.2e-09	51119	! AC090865 Homo sapiens chromo
9b_htg:AC079780	+	305.00	306.35	8.7e-09	165600	! AC079780 Homo sapiens chromo
9b_htg:AC025556	+	298.00	298.63	2.3e-08	159122	! AC025556 Homo sapiens chromo
9b_in:CEM88	+	276.50	288.23	8.9e-08	32723	! Z34802 Caenorhabditis elegans
9b_in:AC016961	+	257.50	249.66	1.2e-05	202643	! AC016961 Homo sapiens chromo
9b_in:DMUB	+	248.50	274.60	5.1e-07	4203	! X99340 D.melanogaster genomic
9b_in:AC016961	+	242.50	232.34	0.0001	202643	! AC016961 Homo sapiens chromo

```

211..:1990
/function="putative RNA binding protein"
/note="Kh-domain containing protein; overexpressed in
cancer"
/codon_start=1
/product="Koc1"
/protein_id="AAD09223.1"
/db_xref="GI:4098297"
/translation="MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTYAFVDC
PDESWALKAIEALSGKIEVHESVPKRRIRKIQIRNIPHLQWEVLDLSLV
QXGVESCEQVNTDSEYAVVNTYSSKQAOARALDKLNGFOLNFTKAVYIPDEMAA
QONQLOPGRGLGCGSRGSGSGSKQKPCDLPRLPLVPTQFVCAITGEGKATY
RNIITQKQSDIVHRKKNAGAAEKSTITLSTPEGTSAAACKSILELHHREAQDIKTTEE
IPKLILAHNNFVRLGIGKGRNLKKEQDITDKITISPLQELTINPENTITVKGVE
TCAKABEEELKTIRESYENDIAISMNLQIATLPGLNALNGLFPFSGMPPTSGPPSA
MVPVYPPQSEETSTVHQIPALSGAITGQKQIKQLSRFAGASIKIAPAEAPDPAK
TRRWITFQPEAQFKAQRIYGIKEENFVSPKEEVKLEAHIRVPSFAAGRVIRGGKK
TYNELQNLSSAEVVVPDPTDENQVNVVKTCTHFYACQVAQRKIQEILTQVQHQQQ
KALQSGPQSRRK"
1277 a 830 c 851 g 1181 t 16 others
BASE COUNT
ORIGIN

```

Quality:	2956.00	Length:	579
Ratio:	5.105	Gaps:	0

US-09-685-696-176 x HSU76705 ..
Align seg 1/1 to: Hsu76705 from: 1 to: 4155

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
 |||||
 251 ATGAACAAACTGTATATCGGAAACCTCAGCGAGAACGCCGCCCTCGGA 300

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||

301 CCTAGAAAGTATCTTCAAGAGCGCCACAGATCCCGGTGTCGGGACCCCTTCC 350

34 euVallysthrGlytyrAlapheValaspCysProaspGluSerTrpAla 50
|||||

```

351 TGGTGAAGACTGGCTACGGTTCTGGTGGACTGCCCGGACGAGAGCTGGGCC 400
51  LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
401 CTCAAGGCGCATCGAGGCGCTTTTCAGGTAAATAGAACTGCACGGGAAACC 450
67  oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeu 84
451 CATAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAAAGGATTCGGAACCTC 500
84  InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
501 AGATACGAATATCCCGCTCATTTACAGTGGAGGTGCTGGATAGTTTA 550
101 LeuValGlnTrpGlyValIleSerCysGluGlnValAsnThrAspSe 117
551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
601 GGAACACTGCAGTTGTAATGTAACTTATTCAGTAAGGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
651 AAGCACTAGACAACTGAATGGATTTACGTTAGAGAAATTTACCTTGAAA 700
151 ValAlaTrpIleProAspGluMetAlaAlaGlnGlnAsnProLeuGln 167
701 GTAGCTATATCCCTGTATGAATGGCCGCCCCAGCAAAACCCCTTGCAGCA 750
167 nProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGly 184
751 GCCCGAGGTCGCGGGGGCTTGGGCGAGAGGGGCTCTCAAGGCGAGGGGT 800
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
801 CTCAGGATCCGTATCCAAAGCAGAAACCATGTGATTTGCCCTCGCGCCGT 850
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
851 CTGGTTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 900
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
901 CATTGGGAACATCAACCAAGAGCCAGCTCTTAAATCGATGCCACCGTA 950
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
951 AAGAAATCGCGGGCTGCTGAGAAAGTCGATTACTATCCTCTCTACTCCT 1000
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
1001 GAAGGCACCTCTCGGGCTTGTAAAGTCTATTCTGGAGATTATGCATAAGGA 1050
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
1051 AGCTCAAGATATAAATTCAGAGAAGAGATCCCTTGAAGATTTTAGCTC 1100
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
1101 ATAATACTTTGTTGGAGCTTATTGGTAAAGGAAGGAATACTTAA 1150
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
1151 AAAATTGAGCAAGACACAGACACTTAAATCAGGATATCTCCATTGCAGGA 1200
317 uLeuThrLeuTrpAsnProGluArgThrIleThrValLysGlyAsnValG 334
1201 ATTGACGCTGTATAATCCAGAACGCACTATTACAGTTTAAAGGCAATGTT 1250
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
1251 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGGAGTCT 1300

```

... 3

```

351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
1301 TATGAAAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATTCCTGG 1350
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
1351 ATTAATCTGAACGCTTGGGTCTGTCCACCCACTTCAGGGATGCCAC 1400
384 toProThrSerGlyProProSerAlaMetThrProProTyrProGlnP 400
1401 CTCCACCTCAGGGCCCTTCAGCCATGACTCTCCCTACCCGCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
1451 GAGCAATCAGAAACGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgP 434
1501 CGGTGCCATCATCGGCAAGCAGGCGCCAGCACATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
1551 TTGCTGGAGCTTCAATTAAAGATTGCTCCAGCGAAGCAGCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGl 467
1601 CTGAGGATGGTGATTATCCTTGGACACACAGAGGCTCAGTTTCAAGGCTCA 1650
467 nGlyArgIleThrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 GGGAGAATTTATGGAAAAATTAAGAAAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
1701 AAGAGTGAAACTTGAAGCTCATATCAGAGTGCCATCTCTTCTGCTGCG 1750
501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAAAAGGAGGCAAAACGGTGAATGAATTCAGAAATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
1801 AAGTGCAGAAAGTTGTTGCTCCCTCGTCCGACGACACCTGATGAGAAATGACC 1850
534 InValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1851 AAGTGGTTGTCAAAATAACTGCTGCTCTCTTCTGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
1901 AGAAAAATTCAGGAAATTTCTGACTCAGGTAAAGCAGCACCACCAACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
1951 GCCTCTGCAAAAGTGACCACTCAGTCAAGACGGGAG 1987

```

seq_name: gb_pr:HSU97188

seq_documentation_block:

LOCUS HSU97188 4181 bp mRNA PRI 11-SEP-1998
 DEFINITION Homo sapiens putative RNA binding protein KOC (koc) mRNA, complete cds.

ACCESSION U97188

VERSION U97188.1 GI:2105468

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4181)

REFERENCE

AUTHORS Mueller-Pillasch,F., Lacher,U., Wallrapp,C., Micha,A., Zimmerhackl,F., Hameister,H., Varga,G., Fries,H., Buchler,M.,

Beger,H.G., Vila,M.R., Adler,G. and Gress,T.M.
 Cloning of a gene highly overexpressed in cancer coding for a novel
 KH-domain containing protein
 Oncogene 14 (22), 2729-2733 (1997)
 97322117
 REFERENCE 2 (bases 1 to 4181)
 Mueller-Pillasch,F., Lacher,U., Wallrapp,C., Micha,A.,
 Zimmerhackl,F., Hameister,H., Varga,G., Friess,H., Buchler,M.,
 Beger,H.G., Vila,M.R., Adler,G. and Gress,T.M.
 Direct Submission
 Submitted (11-APR-1997) Medizinische Klinik, Internal Medicine I,
 Robert-Koch-Str.8, Ulm 89081, Germany
 FEATURES
 Location/Qualifiers
 1. .4181
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="p11.5"
 /cell_type="pancreatic cancer"
 /cell_line="Patu8988t"
 1. .4181
 /gene="koc"
 251. .1990
 /gene="koc"
 /function="possible proliferation of cancer tissues"
 /note="KH-domain containing protein overexpressed in
 cancer"
 /codon_start=1
 /product="putative RNA binding protein KOC"
 /protein_id="AAC35208.1"
 /db_xref="gi:2105469"
 /translation="MNKLIYIGNLSENAAPSDESIFDKAKIPVSGPFLVKTYGAFVDC
 PDGWSWALKEIALSGKTELHGKPIEVESHVPKROIRKIQIRNPPIHLOWEVLDSLLV
 QYGEVCEQVNTDSEAVNVVTVSSKDOARQDLKNGFOLENTFLKVAVIPDEMAA
 QONLQPRGRGLGQRGSSRQSGPSVSKQKPCDPLRLVLPVTFQVGAIGKEGATI
 RNITKQSKIDVHRKENAGAEEKSITILSTPEGTSACKSILEIMHKEAODIKFTEE
 IPLKILAHNNVFLGKEGKRNKKIEQDITKITISPIQELTYLNPERTTVKGNVE
 TCAKEEIEIMKIRIESYENDIASMNLQALHLPGLNALGLFPPTSGMPPTSGPPSA
 MTPYPOFEQSETVTHQFIPLASVGAIGKGOHIIKOLSRFAGASIKIAPEAPDPA
 VMVLIITGPPAQPKAGRIYKIKIENFVSPKEVKLEAHIRVPSFAGRVIGKCK
 TVNELNLSRSEVVPVPRDQTPDENQVVKITGHFYACQVAQRKIQEILTQVKHQHQQ
 KALQSGPPQSRKK"
 1303 a 830 c 851 g 1181 t 16 others

source

gene

CDS

alignment_scores:

Quality: 2956.00 Length: 579
 Ratio: 5.105 Gaps: 0
 Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x HSU97188 ..

Align seg 1/1 to: HSU97188 from: 1 to: 4181

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
 |||||
 251 ATGACAAACTGTATATCGGAACCTTCAGCGAGAACCGCCCTCGGA 300
 |||||
 17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
 |||||
 301 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCTTC 350
 |||||
 34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
 |||||
 351 TGCTGAAGACTGCTACGCGTTCGTGGACTGCCCGGACGAGAGCTGGCC 400
 |||||
 51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
 |||||
 401 CTCAGCCCATCGAGCGCTTTTCAGGTAAATAAGAACTGCGGGAAACC 450
 |||||
 67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
 |||||

|||||
 451 CATAGAAGTTGACACTCGCTCCCAAAAGCAAGGATTTCGGAACCTTC 500
 |||||
 84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
 |||||
 501 AGATACGAATAATCCCGCCCTATTACAGTGGGAGGTGCTGGATAGTTA 550
 |||||
 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
 |||||
 551 CTAGTCCAGTATGAGTGGTGGAGAGCTGTGACCAAGTGAACACTGACTC 600
 |||||
 117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
 |||||
 601 GGAACCTGCAGTTCTAAATGTAACCTATTCCAGTAAAGCAAGCTAGAC 650
 |||||
 134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
 |||||
 651 AAGCACTAGACAAACTGAATGGATTTTCAGTTAGAGAATTTTCACCTTAAA 700
 |||||
 151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
 |||||
 701 GTAGCCTATATCCTGATGAATGGCGCCCGCCAGCAAAACCCCTTGCAGCA 750
 |||||
 167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
 |||||
 751 GCCCCGAGGTGCGCGGGGCTTGGCAGAGGGGCTCTCAAGGCAGGGGT 800
 |||||
 184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
 |||||
 801 CTCAGGATCCGTATCCAAGCAAAACCATGTGATTTGCTCTGCGCCTG 850
 |||||
 201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
 |||||
 851 CTGGTTCACCAACCAATTTGTTGGAGCCATCATAGGAAAAAGAGGTGCCAC 900
 |||||
 217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
 |||||
 901 CATTCGGAACATCACCACCAACAGACCCAGTCTAAATCGATGTCCACCGTA 950
 |||||
 234 yGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
 |||||
 951 AAAAAATGCGGGGCTGCTGAGAAGTCGATTACTATCTCTCTACTCCT 1000
 |||||
 251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGl 267
 |||||
 1001 GAAGGCACCTCTCGGCTTGTAGTCTATTCTGGAGATTATGCATAAGGA 1050
 |||||
 267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
 |||||
 1051 AGCTCAAGATATAAAATTACAGAAAGATCCCTTGAAGATTTTAGCTC 1100
 |||||
 284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
 |||||
 1101 ATAATAACTTTTGTGGACGCTCTATTGGTAAAGAAAGAAATCTTTAAA 1150
 |||||
 301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGl 317
 |||||
 1151 AAAATTGAGCAAGACACAGACACTAAATACGATATCTCCATTGCGAGGA 1200
 |||||
 317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
 |||||
 1201 ATTGACGCTGTATAATCCAGAACGCACATATTACAGTTAAAGCAATGTTG 1250
 |||||
 334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
 |||||
 1251 AGACATGTGCCAAAGCTGAGGAGAGATCATGAAGAAAAATCAGGGAGTCT 1300
 |||||
 351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
 |||||
 1301 TATGAAATGATATGCTTCTATGAATCTTCAAGCACATTTAATTCCTGG 1350
 |||||
 367 yLeuAsnLeuAlaLeuGlyLeuPheProThrSerGlyMetProp 384
 |||||

1351 ATTAATCTGAACGCTTGGGTCTGTTCCACCCCACTTCAGGGATGCCAC 1400
384 roProThrSerGlyProProSerAlaMetThrProProTyProGlnPhe 400
1401 CTCCACCTCAGGGCCCTTCAGCCATGACTCTCCCTACCCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
1451 GAGCAATCAGAAACGAGAGCTGTTTCATCAGTTTATCCAGAGCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyHisIleLysGlnLeuSerArgp 434
1501 CGTCCCATCATCGCAGCAGGCGCCACACATCAAGCAGCTTCTCCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
1551 TTGCTGGAGCTTCAATTAGATTGCTCCAGCGGAGCACCAGATGCTTAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGl 467
1601 GTGAGGATGGTGATTATCCTGACGACCAAGAGGCTCAGTTCAGGCTCA 1650
467 nGlyArgIleThrGlyLysIleLysGluGluAsnPheValSerProLysG 484
1651 GGAAGAATTTATGGAAAATTTAAGAAAGAAACTTTGTTAGTCTTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
1701 AAGAGTGAACCTTGAACCTCATATCAGAGTGCCATCTTGTGCTGCGC 1750
501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
1751 AGAGTTATTGGAAAAGGAGGCAAAACGGTGAATGAACCTTCAGAAATTTGTC 1800
517 rSerAlaGluValValValProArgAspGlnThrProAspGluAsnAspG 534
1801 AAGTGCAGAGTGTGTGTCCTCGTACGACAGACCTGATGAGATGACC 1850
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
1851 AAGTGTGTCAAAATACTGGTCACITCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
1901 AGAAAAATTACAGAAATTTCTGACTCAGGTAAGCAGCACCACCAACAGAA 1950
567 salAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
1951 GGCTCTGCAAGTGGACCACCTCAGTCAAGACGGAAG 1987

seq_name: gb_pr:AF117108

seq_documentation_block:
LOCUS AF117108 4171 bp mRNA PRI 26-JAN-1999
DEFINITION Homo sapiens IGF-II mRNA-binding protein 3 (IMP-3) mRNA, complete cds.

ACCESSION AF117108
VERSION AF117108.1 GI:4191611

KEYWORDS

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4171)

AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,

Wewer,U.M. and Nielsen,F.C.

TITLE A family of insulin-like growth factor II mRNA-binding proteins

represses translation in late development

JOURNAL Mol. Cell. Biol. 19 (2), 1262-1270 (1999)

REFERENCE 2 (bases 1 to 4171)

AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,

Wewer,U.M. and Nielsen,F.C.

TITLE Direct Submission
JOURNAL Submitted (30-DEC-1998) Institute of Molecular Biology, University of Copenhagen, Soelvgade 83H, Copenhagen DK-1307, Denmark
FEATURES
Location/Qualifiers
source 1..4171 /organism="Homo sapiens"
gene 1..4171 /db_xref="taxon:9606"
CDS 267..2006 /gene="IMP-3"
/note="binds multiple sites in IGF-II mRNA 5'UTR; identical to KH-domain containing protein overexpressed in cancer (koc)"
/codon_start=1
/product="IGF-II mRNA-binding protein 3"
/protein_id="AAD09828.1"
/db_xref="GI:4191612"
/translation="MNKLYIGNSENAAPSDESIFPKDAKIPVSGPPLVKVTGVAFVDC PDESWALKATEALSNGRIEHLGKPIEVEHSVPKRIRKLIQIRNIPHLQWELDSLIV OYGVESCEQVNTDSTAVVNTYSSKQARQALDKLNGFQLENFTLKVAY IPDEMAA QONPQQPRGRLGORGSSROGSPGVSQKQPCDLPLRLVPTQFVQATIGKEGATI RNITKQTSKIDVHRKENAGAEKSIITLSTPEGTSNACKSILEIMHKEAQDIFKTEE IPLKILAHNNFMKRTIESYENDIASMNLOAHLIPGLNINLALGFLPPTSGMPPTSGPPSA MTPPYPQFEQSETETVHLFTIPALSVGAIIGKQGHKQLSRFAGASIKIAPABAPDAK VRMVIITGPPEAQFKAQRIYKIKIENEFVSPKEVKLEAHIRVPFAAGRVIGKGGK TVNELONLSSAEVVVPDQTPDENDQVVKVITGHFYACQVAQRKIQLIETLVQVKHQOQ KALQSGPPQSRKK"
BASE COUNT 1292 a 822 c 853 g 1199 t 5 others
ORIGIN
alignment_scores:
Quality: 2949.00 Length: 579
Ratio: 5.102 Gaps: 0
Percent Similarity: 99.827 Percent Identity: 99.827
alignment_block:
US-09-685-696-176 x AF117108 ..
Align seg 1/1 to: AF117108 from: 1 to: 4171
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
267 ATGAACAACTGTATATCGAAACCTCAGCGAGAACGCGCCCTCCGGA 316
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
317 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCCTTC 366
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
367 TGGTGAAGACTGGCTACCGTTCGTGGACTCCCGGACGAGCTGGGCC 416
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
417 CTCAAGGCCATCGAGGCGCTTTCAGGTAATAATAGAACTGCACGGGAACC 466
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
467 CATAGAAGTTGAGCACTCGGTCCCAAAAAGGCAAGGATTCGGAACATTC 516
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
517 AGATACGAATAATATCCCGCTCATTTACAGTGGAGGTGCTGGATAGTTTA 566
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
567 CTAGTCCAGTATGAGTGTGTGAGAGCTGTGAGCAAGTGAACACTGACTC 616
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||


```
617 GGAACCTGCAGTTGTAATGTAACCTATTCCAGTAAGGACCAAGCTAGAC 666
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
667 AAGCAGCTAGACAACTGAATGGATTTCAGTTAGAGAATTTCCACTTGAAA 716
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
717 GTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGACGAC 766
167 nProArgGlyValArgArgGlyLeuGlyGlnArgGlySerSerArgGlnG 184
|||||
767 GCCCGAGAGTCGCCGGGGCTTGGCGAGAGGGCTCCTCAAGCGAGGGT 816
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
817 CTCCAGGATCCGTATCCAAAGCAGAAACCATGTGATTTGCCTCTGCGC 866
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
|||||
867 CTGGTTCCCAACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCC 916
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArg 234
|||||
917 CATTCGGAACATCACCAACAGACCCAGTCTAAATCGATGTCACCGTA 966
234 ysGluAsnAlaGlyAlaAlaGlyLysSerIleThrIleLeuSerThrPro 250
|||||
967 AAGAAATGCGGGGCTGCTGAGAAGTCGATTACTATCCTCTACTCCT 1016
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
|||||
1017 GAAGGCACCTCTCGGGCTTGTAAGTCTATTCTGGAGATTATGCATAAG 1066
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAla 284
|||||
1067 AGCTCAAGATATAAAATTCACAGAAGAGATCCCTTGAAGATTTAGCT 1116
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
1117 ATAATACTTTGTTGGAGCTCTATTATGTTAAAGAGGAAGAAATCTTAA 1166
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
|||||
1167 AAAATTGAGCAAGACACAGACACTAAAATCAGATATCTCCATTGCAG 1216
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
1217 ATTGACGCTGTAAATCCAGAAGCCACTATTACAGTTTAAAGGCAATGT 1266
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1267 AGACATGTGCCAAGCTGAGGAGAGATCATGAAGAAATCAGGAGTCT 1316
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
|||||
1317 TATCAAAATGATATTGCTTCTATCAATCTTCAAGCACATTTAATCTCG 1366
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1367 ATTTAAATCTGAACCCCTTGGCTGTGTTCCCAACCCACTTCAGGGATG 1416
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1417 CTCCCACTCAGGGCCCTTTCAGCCATGACTCTCTCCCTACCCGAGTTT 1466
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1467 GAGCAATCAGAAACGAGAGCTGTTTCATCTGTTTATCCAGCTCTATCAG 1516
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArg 434
|||||
1517 CGGTGCCATCATCGCAAGCGAGGGCCAGCATCAAGCAGCTTCTCGCT 1566
```

```
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1567 TTGCTGCAGCTTCAATTAAAGATTGCTCCAGCAGAAGCACCAGATGCTAAA 1616
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG 467
|||||
1617 GTGAGGATGGTATATCACTGGACCACCAGAGGCTCAGTTCAAGGCTCA 1666
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLys 484
|||||
1667 GGGAGAAATTTATGAAAAATTAAGAAGAAACTTTGTTAGTCTTAAAG 1716
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1717 AAGAGGTGAAACTTGAAGTCTATATCAGAGTGCATCTCTTGTGCTGCTG 1766
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1767 AGAGTTATTGAAAAAGAGGCAAAACGGTGAATGAATTCAGAAATTTGTC 1816
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1817 AAGTGCAGAAAGTTGTTGCCCTCGTGACACACACCTGATGAGATGACC 1866
534 InValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1867 AAGTGGTTGTCAAAATAACTGGTCACCTCTATGCTTGCACAGTTGCCCA 1916
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
|||||
1917 AGAAAAATTCAGGAATTTCTGACTCAGGTAAGCAGCACCACCAACAGAA 1966
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1967 GGCTCTCAAGTGGACACCTCAGTCAAGACGGAAG 2003
seq_name: gb_ro:AB046173
seq_documentation_block:
LOCUS AB046173 1740 bp mRNA ROD 23-DEC-2000
DEFINITION Mus musculus mImp3 mRNA for igf2 mRNA-binding protein 3, complete
cds.
ACCESSION AB046173
VERSION AB046173.1 GI:119333383
KEYWORDS
SOURCE Mus musculus cDNA to mRNA.
ORGANISM
REFERENCE
AUTHORS Mori,H., Sakakibara,S., Imai,T., Nakamura,Y., Iijima,T., Suzuki,A.,
Yuasa,Y., Takeda,M. and Okano,H.
TITLE Expression of mouse igf2 mRNA-binding protein 3 and its
implications for the developing central nervous system
JOURNAL J. Neurosci. Res. (2001) In press
REFERENCE
AUTHORS Mori,H., Okano,H., Sakakibara,S. and Imai,T.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) Hiroshi Mori, Osaka Univ. Graduate School
of Medicine, Division of Neuroanatomy (D12); 2-2 Yamadaoka, Suita,
Osaka 565-0871, Japan (E-mail:hmori@nana.med.osaka-u.ac.jp,
Tel:81-6-6879-3581, Fax:81-6-6879-3589)
FEATURES
Source
1..1740
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..1740
/genes="mimp3"
CDS
1..1740
/genes="mimp3"
/codon_start=1
/product="igf2 mRNA-binding protein 3"
```

```

/protein_id="BAB19755.1"
/db_xref="GI:11933384"
/translation="MNKLYIGNLSDHAGPADLESFVKDAKIPVAGPFLVKGTGYAFVDC
PDEGWAIAEALSGKMLHGKPMVEHVSVPKRRIRKIQIRNIPPHLOWEVLDSLLV
QYGVESCEQVNTDSTAVVNTYSSKDAQALDKLNGFQLENFLLKWAYIPDETA
QNPSPQLKRRGPGGSGRSRASPQSVSKQPCDPLRLVPTQVGAIIQKREGATI
RNITKQTSKIDVHRKENTGAAPKSTILSTPEGTSAAKSILEIMHKBAQDIKFTEE
IPKILAHNFVRLTGKGRNLKITEQDTRKIIISPLQELTLYNPERTITVKGSGVE
TCABAEEMKKIKRESYENDIASMNLQAHILPGLNLAGLFPPTSGMPPTSGPPST
LTPYPQFQSETEVHLFIPALSYGAIIGKOGHILKOLSRFAGASIKITAPAEAPDAK
VMVLIITGPEPAQKQRIYGIKKEENFVSPKEEVKLEAHIRVSPFAAGRVIGKGGK
TVNELQSLSRAEYVPRDQTPDENQDVVKITGHFYACQVQARKIQEILTVQKHQQQ
KALQSGPPQSRKK"
BASE COUNT      500 a   431 c   442 g   367 t
ORIGIN

```

alignment_scores:

Quality:	2858.00	Length:	579
Ratio:	4.988	Gaps:	0
Percent Similarity:	98.964	Percent Identity:	96.373

alignment_block:

US-09-685-696-176 x AB046173 ..

Align seg 1/1 to: AB046173 from: 1 to: 1740

```

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaLaProSerAs 17
|||||
1 ATGAACAAATTTGATCATCGGAACTTCAGGCACCGCCGCCCGCGGA 50
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
51 CCTGGAAGTCTCTCAAGGACGCTAAGATCCCGTGGCGGCCCTTCC 100
|||||
34 euValLysThrGlyTyrAlaPheValaspCysProaspGluSerTrpAla 50
|||||
101 TGGTGAAGACGGGTACGCTTCGTGGACTGCCCGGACGAGGCGTGGCC 150
|||||
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 CTCAGGCCATCGAGGCGCTTCAGGTAAATGGAACATACATGGAACCC 200
|||||
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 GATGGAAGTTGAGCACTCGGTCTTAAACGCGCAGAGGATTCGTAACATC 250
|||||
84 lnileArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATACGAAATATCCGCCCCACCTTACAATGGGAGGTGCTGGATAGTTA 300
|||||
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 CTAGTCAGATGAGTGGTGGAGAGCTGTGAGCAAGTGAACACGGATTC 350
|||||
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
351 GGAACGCGCAGTTGTAATGTAACCTATTCCAGTAAGGACCAAGCTAGAC 400
|||||
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCCTTAGACAAACTGAATGGATTCCAGTTAGAGAACTTCACCTTGAAG 450
|||||
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||
451 GTTCCTCATCTCCAGATGAACATGCTGCCACGCAAAATCCCTCGCCACA 500
|||||
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
|||||
501 GCTCCGGGGCGCGGGCGGCGAGCGGGCGGCTCATCCAGGCGAGCGGT 550
|||||
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||

```

```

551 CTCAGGATCGGTGTCCAAAGCAGAAACCCCTGTGACCTGCCACTGCCCTG 600
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
|||||
601 CTGGTTCACCACCGATTGTTGGAGCCATTATAGGAAAGAGGTGCTTAC 650
|||||
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
|||||
651 CATTCGCAACATCAAAAGCAGACCCAGCTCTAAAATCGATGTCATCCTG 700
|||||
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
|||||
701 AGGAGATACAGGGCGCGGAGAGTCCATTACTATCTCTTACCCT 750
|||||
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGl 267
|||||
751 GAAGCACCTCTGCAGCTTGTAACTTATCTGGAGATTATGCATAAGA 800
|||||
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
801 AGCTCAAGATATAAAATTCAGAGGAGATTCCCTTGAAGATCTTAGTCT 850
|||||
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
851 ACAATAACTTTGTAGCGCGTCTCATTTGTTAAAGAGGAAGAAACCTTAA 900
|||||
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGl 317
|||||
901 AAAATCGAGCAAGCACGACACATAAAATCACATATCTCCATTCACAGA 950
|||||
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
951 ACTGACGCTGTACAATCGGAACGACCATTTACAGTGAAGGACGAGTGTG 1000
|||||
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1001 AGAGTGTGCCAAGCGGAGGAGAAATAATGAAGAGATCAGGAGTCT 1050
|||||
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
|||||
1051 TATGAAATGATATTGCTTCATGAACTTCAAGCACATTTAATCCCTGG 1100
|||||
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1101 ATTTAAATCTGAATGCTTGGTCTGTCCACCCACGTCAGGAGTGCCAC 1150
|||||
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1151 CTCCACCTCAGGGCCCCCTTCAACCTTGACTCCTCCCTACCCACAATTT 1200
|||||
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1201 GAGCAATCAGAGCGGAGACTGTGCATCTGTTATTCCCGCCCTGTCGT 1250
|||||
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
|||||
1251 TGGCGCCATCATTTGGCAAGCAGGCGCAACACATCAACAGCTTTCTCGCT 1300
|||||
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1301 TTGGGGAGCTTCGATTAAAGTCTCCAGCCGAGGACCCAGATGCTTAA 1350
|||||
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaGl 467
|||||
1351 GTGCGGATGGTATATACATGGACCACCCAGAGGCTCAGTTCAAGGCTCA 1400
|||||
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1401 GGGGAAGAATTTATGGAAAAATTAAAGAAAGAAACTTTTGTAGTCTTAAAG 1450
|||||
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1451 AAGAGGTGAAACTTTGAAGCTCACATCAGAGTGGCGTCTCTTTTGTCTGGC 1500

```

501 ArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1501 AGAGTTATTGGGAAGGAGGCAAAACGGTGAATGAGCTCCAGAGTTATC 1550
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1551 AAGTGTGAAGTTGCTGCTCCCGGACACGACACTGATGAGATGATC 1600
534 InValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1601 AAGTAGTTGTCAAATAACTGCCACTTCTATGCTGCCAGGTGCCAG 1650
551 ArgLysIleGlnGluLeuThrGlnValLysGlnHisGlnGlnGlnIly 567
|||||
1651 AGGAAATTCAGGAATTCGACTCAGGTAAAGCAGCAGCAGCAGAGAA 1700
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1701 AGCTTGCAGAGTGACACCTCAGTCAAGCGCGAAG 1737

seq_name: gb_pr:HS497J21

seq_documentation_block:

LOCUS HS497J21 129432 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 497J21 on chromosome 6q26-27.
Contains a KOC (KH-domain containing transcript overexpressed in
cancer) pseudogene, genomic marker D6S193, ESTs, STSS and GSSs, and
a ca repeat polymorphism, complete sequence.

ACCESSION AL023775
VERSION AL023775.1 GI:3646096
KEYWORDS HTG; ca repeat polymorphism; D6S193; KH-domain; KOC.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 129432)
Tubby.B.

TITLE Direct Submission
AUTHORS Submitted (17-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
JOURNAL requests: clonerequest@sanger.ac.uk

COMMENT On Sep 24, 1998 this sequence version replaced gi:3550063.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 497J21.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 427A4 (298049) is at 129333 in this
sequence.

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

Assembly was confirmed by BamHI and HindIII digests. Discrepancy
noted in EcoRI digest.

This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>

497J21 is from the library RPC13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

Location/Qualifiers

1..129432

/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
/chromosome="6"
/map="Q26-27"
/clone="RP3-497J21"
/clone_lib="RPC1-3"
complement(1..62)
/note="AluSx repeat: matches 62. .1 of consensus"
27..424
/note="match: STS AL022447"
63..231
/note="MER20 repeat: matches 53. .218 of consensus"
433..591
/note="3 copies 53 mer 86% conserved"
complement(1513..1915)
/note="MLT1B repeat: matches 382. .2 of consensus"
complement(2671..2713)
/note="MLT1B repeat: matches 47. .3 of consensus"
complement(3291..3586)
/note="AluSx repeat: matches 298. .1 of consensus"
4989..5126
/note="L1MB6 repeat: matches 1. .141 of consensus"
5137..5648
/note="L1MB5 repeat: matches 382. .923 of consensus"
complement(7362..7463)
/note="MIR repeat: matches 194. .87 of consensus"
8372..8668
/note="AluSx repeat: matches 2. .300 of consensus"
8670..8703
/note="17 copies 2 mer aa 82% conserved"
14252..14295
/note="22 copies 2 mer gt 98% conserved"
14254..14301
/note="3 copies 16 mer 94% conserved"
14254..14297
/note="11 copies 4 mer gtgt 98% conserved"
14441..14481
/note="MIR2 repeat: matches 99. .139 of consensus"
complement(16259..16324)
/note="MIR repeat: matches 157. .92 of consensus"
complement(18532..18711)
/note="MIR repeat: matches 218. .29 of consensus"
20727..21002
/note="AluSx repeat: matches 2. .296 of consensus"
22965..23004
/note="10 copies 4 mer caca 83% conserved"
complement(24989..25171)
/note="L1MB3 repeat: matches 931. .738 of consensus"
complement(25204..25504)
/note="AluSp repeat: matches 303. .1 of consensus"
complement(25761..26017)
/note="AluY repeat: matches 301. .45 of consensus"
complement(27102..27194)
/note="MIR repeat: matches 142. .50 of consensus"
30090..30204
/note="5 copies 23 mer 94% conserved"
complement(31383..31419)
/note="MIR2 repeat: matches 113. .76 of consensus"
31912..32368
/note="match: STS S53188; part 1 of genomic marker D6S193"
32276..33267
/note="31 copies 32 mer 91% conserved"
33141
/note="tandem repeat"
33204
/note="tandem repeat"
33222..33396
/note="tandem repeat and single clone region"
33268..33555
/note="9 copies 32 mer 88% conserved"
33268..33539
/note="17 copies 16 mer 89% conserved"
33473..33793
/note="match: STS S53190; part 2 of genomic marker D6S193"

repeat_region	complement(34751..35259) /note="MER21B repeat: matches 754..240 of consensus"
repeat_region	complement(35356..35625) /note="MER21B repeat: matches 282..5 of consensus"
repeat_region	36446..36618 /note="MLR1B repeat: matches 1..179 of consensus"
repeat_region	36619..36921 /note="Alusq repeat: matches 1..303 of consensus"
repeat_region	36927..37136 /note="MLR1B repeat: matches 175..390 of consensus"
repeat_region	38179..38310 /note="2 copies 66 mer 92% conserved"
repeat_region	38187..38318 /note="4 copies 33 mer 91% conserved"
repeat_region	complement(39918..40225) /note="Alusq repeat: matches 302..1 of consensus"
repeat_region	complement(41241..41409) /note="L1 repeat: matches 5390..5208 of consensus"
repeat_region	42029..42091 /note="MR2 repeat: matches 69..138 of consensus"
misc_feature	complement(43307..43755) /note="match: GSS B76271"
repeat_region	46203..46713 /note="MER1A repeat: matches 1..527 of consensus"
repeat_region	complement(50808..51108) /note="Alusq repeat: matches 303..2 of consensus"
repeat_region	52433..52723 /note="Aluwb repeat: matches 5..301 of consensus"
repeat_region	53862..54166 /note="MER2 repeat: matches 1..324 of consensus"
repeat_region	54319..54430 /note="7 copies 16 mer 81% conserved"
repeat_region	54322..54417 /note="3 copies 32 mer 89% conserved"
repeat_region	54720..54767 /note="24 copies 2 mer tg 88% conserved"
repeat_region	58071..58370 /note="Alusq repeat: matches 1..301 of consensus"
repeat_region	58587..58738 /note="4 copies 38 mer 86% conserved"
misc_feature	complement(59963..60560) /note="match: GSS Aq001508"
repeat_region	60205..60273 /note="3 copies 23 mer 84% conserved"
repeat_region	61253..61299 /note="MR repeat: matches 104..150 of consensus"
repeat_region	63923..64224 /note="Alusq repeat: matches 1..303 of consensus"
repeat_region	complement(65464..65556) /note="MER7A repeat: matches 335..243 of consensus"
repeat_region	65556..65632 /note="MER6 repeat: matches 1..77 of consensus"
repeat_region	65631..65733 /note="MER6 repeat: matches 248..346 of consensus"
repeat_region	65730..65984 /note="MER6 repeat: matches 603..864 of consensus"
repeat_region	complement(66013..66314) /note="Alusq repeat: matches 302..1 of consensus"
repeat_region	complement(66318..66528) /note="MER7A repeat: matches 242..1 of consensus"
repeat_region	complement(67223..67664) /note="MLR1C repeat: matches 465..3 of consensus"
repeat_region	complement(69999..70266) /note="L1 repeat: matches 3386..3129 of consensus"
repeat_region	70267..70292 /note="13 copies 2 mer tc 92% conserved"
repeat_region	71520..71625 /note="2 copies 53 mer 85% conserved"
repeat_region	complement(72592..72724) /note="Aluao repeat: matches 134..1 of consensus"
repeat_region	74027..74329 /note="AluJ repeat: matches 1..302 of consensus"
repeat_region	complement(75896..76260) /note="AluJ repeat: matches 1..302 of consensus"

```
79931 |TGTCCACCGCTAAAGAAATGACGGGCTGCTGAGAAAGTCATTACTATCC| 79980
247  |euSerThrProGluGluThrSerAlaAlaCysLysSerIleLeuGluIle| 263
79981 |TCCTACTCTGAAGCACCCTGCGGCTGTGAAGTCTATCTGGAGATT| 80030
264  |MetHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu| 280
80031 |ATCGGTAAGGAAACTCAAGATGTAATAATCACAGAAGAGATCCCTTGAA| 80080
280  |sIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGly| 297
80081 |GATTTAGCTCAAAATAACTTTCTGGCGCTCTATTGTTAAAGAAGAA| 80130
297  |rAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSer| 313
80131 |GAATCTTAAANAANTTGACAGACACAGACACTAAATATCAGATATCT| 80180
314  |ProLeuGlnGluLeuThrLeuTyAsnProGluArgThrIleThrVally| 330
80181 |CCATTCGAGGAAGTATGTCAGTATATATCCAGAACACACCATACAGCTAA| 80230
330  |sGlyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysI| 347
80231 |AGCAATGTTGAGACATGTGCCAAGCTGAGGAAGAGATCATGAAGAAA| 80280
347  |leArgGluSerTyGluAsnAspIleAlaSerMetAsnLeu.GlnAlaHi| 363
80281 |TCAGGAGCGCTTAGGAAATGATACATCTCTATG...ATTCAAGCACA| 80327
363  |sLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrS| 380
80328 |TTTAATTCCTGGATTAAATCTGACGCTTGGCTCTCTCCACCCACTT| 80377
380  |erGlyMetProProProThrSerGlyProProSerAlaMetThrProPro| 396
80378 |CAGGATGCCACCTCCACCTCAGGCGCCCTTCAGCCATGACTCTCTCC| 80427
397  |TyProGlnPheGluGlnSerGluThrGluThrValHisGlnPheIlePr| 413
80428 |TCCCGCAGTCTGAGCAATCAGATCCGAGACTGTCATCTGTTATCCC| 80477
413  |oAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysG| 430
80478 |GGCTCTATCATGTCAGACCTCATCAGCAGAGCGGCCACCATCAAGC| 80527
430  |InLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAla| 446
80528 |AGCTTCTCGCTTGTGGAGCTTCAAGTAAGATTGCTCCAGTGAAGCA| 80577
447  |ProAspAlaLysValArgMetValIleIleThrClyProGluAlaGl| 463
80578 |CCAGATCTAAGTGAGGATGGTATGATCGCTGGATCACCAGAGGCTCG| 80627
463  |nPheLysAlaGlnGlyArgIleTyGlyLysIleLysGluGluAsnPheV| 480
80628 |GTTCAAGGCTCAGGAAGAT.TATGGAATAATCAAGAAGAAACTTCG| 80676
480  |alSerProLysGluGluValLysLeuGluAlaHisIleArgValProSer| 496
80677 |TTAGTCTCTAAGAAGAGGTGAAACTTGAAGCTCATATCAGATGCCATCC| 80726
497  |PheAlaAlaGlyArg.ValIleGlyLysGlyGlyLysThrValAsnGluL| 513
80727 |TTGCTCTCGCAGATTACT...GGAAGAAGGAGGCAAAACCGTGAATGAA| 80773
513  |euGlnAsnLeuSerSerAlaGluValValProArgAspGlnThrPro| 529
80774 |TTCAGATTTGTCAGTGCAGAGTGTGTTCTCCCTTGTGACCAGACACT| 80823
530  |AspGluAsnAspGlnValValLysIleThrGlyHisPheTyAlaCy| 546
```

```
80824 GATGAGAATGACCAAGTGGTTGTCAAAATAACTGGTCACCTTCTATGCTTG 80873
546 sGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnH 563
80874 CCAGTTGCCACAGAGAAATTCAGGAATTCGACTCAGTAAAGCAGC 80923
563 isGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
80924 AGCAACAACAGAGGCTCTGCAAAATGGACCACCTCAGTCAAGACGGAAG 80973

seq_name: gb_ov:AF064633

seq_documentation_block:
LOCUS AF064633 2697 bp mRNA VRT 03-JUN-1998
DEFINITION Xenopus laevis Vg1 RNA binding protein variant A mRNA, complete cds.
ACCESSION AF064633
VERSION AF064633.1 GI:3172446
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 2697)
AUTHORS Hawin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman
Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE RNA binding protein conserved in both microtubule and
microfilament-based RNA localization
JOURNAL Genes Dev. (1998) In press
REFERENCE 2 (bases 1 to 2697)
AUTHORS Hawin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman
Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Biochemistry, Cambridge University, Tennis
COURT Road, Cambridge CB2 1GA, UK
FEATURES
source
1..2697
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="oocyte"
299..2080
/note="Vg1-RBP; Vg1 localization element binding protein"
/codon_start=1
/product="Vg1 RNA binding protein variant A"
/protein_id="AAC18597.1"
/db_xref="GI:3172447"
/translation="MNKLYIGNLSENVSPPDLESFLKESKIPFTGQFLVKSGYAFVDC
PDETWMKAIDTLSGVELHGKVIIEVHSVPKRQRSLQIRNIPPHLOWEVLDSLLA
QYGTVENCEQVNTDSEAVVNVYANKHARQGLKNGYOLENYSLVKTYIPDENAT
POSQSQLOQPOOHQPGRRGPGPARQSGPAAARPKQSEVPLRMLVPTQFVGA
LIGKEGATIRNTKOTQSKIDIRKENAGAAEKPTIHSTPEGCAACKIIMEIMKE
AQDTFTTEIPLKILAHNNFVORLIGKSRNLKIEODPTDKITISPLQDLTLNPER
TITVKGSIETCAKEEENKKTRESYNDIAANLQALHLPGLNLNALGLFPSSSGM
PPPSAGVSPSTTSYSPPPQPESETVHLFIPLAVALGAILIKQGHQHIKOLSRFAGS
IKIAPAEQDPAKLRWVITGPPEAKQAGRIYGLKLEENFPGKEEVEKLEAHIKVPS
YAARGVIGKGTNQLNLTAEEVVPRDQTPDENQDVVVKITGHFYASLAQRKIQ
ETLAQVRRQQQQQKTAQSGQPQPRK"
BASE COUNT 855 a 572 c 588 g 681 t 1 others
ORIGIN

alignment_scores:
Quality: 2483.00 Length: 593
Ratio: 4.515 Gaps: 6
Percent Similarity: 92.749 Percent Identity: 82.125

alignment_block:
US-09-685-696-176 x AF064633
Align seg 1/1 to: AF064633 from: 1 to: 2697
1 MetAsnLysLeuTyIleGlyAsnLeuSerGluAsnAlaLaProSerAs 17
```

|||||
299 ATGACACAGCTGTATATTGGAAACCTTAGCGAAACAGTCAGGCCCCCGGA 348
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
349 TTTGGAAAGTCTCTTCAAGGAGTCGAAGATCCCTTTCACCGGCGAGTTTC 398
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
399 TGGTCAAAAGTGTACACGCTTTGTGGACTGTCCGGATGAGACGTGGGCT 448
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
449 ATGAAGGCATCGACACCTCTCAGGGAAGTGGAGTTGCATGGCAAGT 498
67 oLeGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
499 GATACAAGTTGAACACTCAGTACCCCAAAAGGCAAGAGTCAAGAGCTTC 548
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
549 AGATAAGGAACATACACCTCATCTACAGTGGGAGGTACTGGACAGCCTA 598
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
599 TTAGCACAGTATGGACAGTGGAAACCTGTGAGCAAGTTTAACTGATTC 648
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
649 AGAAACTGCAGTAGTGTAATGAACATATGCCAATAAGGAGCAGCGTAGAC 698
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
699 AAGGACTCGAAAAATTAAACGGCTATCAGCTTGAGAACTATAGCCTGAA 748
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||
749 GTGAGCTATACCTGTGTAATGGCAACACCGCAATCACCATCCAGCA 798
167 n.....ProArgGlyArgArgGlyLeuGlyG 176
|||||
799 GCTTCAGCAGCGCCAGCAGCAACACCTCAAGGGAGGCGTGGGTTGGGC 848
176 InArgGlySerSerArgGlnGlySerProGlySerValSerLysGlnLys 192
|||||
849 AAAGGGACCTGCCAGGAGGCGTCCCTGTGCTGCTGCAAGACCTAAG 898
193 Pro....CysAspLeuProLeuArgLeuLeuValProThrGlnPheValGl 208
|||||
899 CCTCAGTCAGAGGTCCACTGAGGATGTGGTGGCCACACAGTTGTGG 948
208 yAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnT 225
|||||
949 GCGGATCATTTGAAAGAGGGTGGCCACCATCAGGACATCATCTAAACAGA 998
225 hrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGlu 241
|||||
999 CCCAGTCAAAAATAGATATTACCGTAAAGAAAATGCTGGTGCACAG 1048
242 LysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLy 258
|||||
1049 AAACCGATTACGATTTCATCAACCCCTCAAGGCTGTTTCAGCTGCATGCAA 1098
258 sSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrG 275
|||||
1099 GATCATTTATGAGATCATCGCAGAGGAGCTCAGGATACCAAGTTCACGT 1148
275 luGluIleProLeuLysIleLeuAlaHisAsnAsnPheValIcIArgLeu 291
|||||
1149 AAGAAATCCCTTAAAAATATTAGCACATAACAATTTTGTGGCGCTCT 1198
292 IleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThrAspTh 308
|||||

seq_name: gb_ov:AF042353

seq_documentation_block:

1199 ATTGGGAAGAAGGAGGAAACCTTAAGRAAATAGACCAAGACACAGATAC 1248
308 rLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluA 325
|||||
1249 CAAAATCAAAATATCTCCACTACAGGACTTGACACTGTACAATCTCTGAAC 1298
325 rGThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341
|||||
1299 GGACCATTTACAGTAAAGGCGACCATTTGACATGTGCCAAAGCCGAGAA 1348
342 GluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMe 358
|||||
1349 GAAGTCATGAAAAAATTTAGGAGTCCTATGAAATGATATTGCTGCTAT 1398
358 tAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyL 375
|||||
1399 GAATCTACAGGCACATTTGATTCCTGGTTTAAACCTGAATGCTGGGCC 1448
375 euPheProPro....ThrSerGlyMetProProProThrSerGlyProPro 390
|||||
1449 TTTTCCACCATCTCTCGTCAGGAATGCCACCACCATCTGCTGGAGTTTCC 1498
391 SerAlaMetThr....ProProTyrProGlnPhe....GluGlnSerGluTh 405
|||||
1499 TCTCCGACAAACATCTGCTTCTTATCCACATTTGGCGACGCCAGAGTC 1548
405 rGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleIleG 422
|||||
1549 AGAGACTGTTCATCTCTTCATCCAGCTTTAGCTGTGGAGCAATATTG 1598
422 lLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
|||||
1599 GAAAGCAAGGACACACATCAACAACTGTCGCGTTTTCAGGAGCATCT 1648
439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValIl 455
|||||
1649 ATTAGATTGCACCGCAGAGGAGCTGTATGCCAACTAAAGATGGTGTAT 1698
455 eIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrG 472
|||||
1699 CATCACTGGACCTCTCTGAAGCCCAATTTAAGGCTCAAGGAGGATCTATG 1748
472 lLysIleLysGluGluAsnPheValSerProLysGluGluValLysLeu 488
|||||
1749 GTAAACTTAAAGAGAAAACTTCTTTGGACCTTAAAGAGAGAGTGAACCTT 1798
489 GluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLy 505
|||||
1799 GAGGCCACATTAAGTGCCGTCATATGCTGCTGGACGTTTATTGGCAA 1848
505 sGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluValV 522
|||||
1849 AGGAGCAAAACAGTAATGAATTCAGATTTAACAAGTCAGAAGTTG 1898
522 alValProArgAspGlnThrProAspGluAsnAspGlnValValLys 538
|||||
1899 TTGTGCCCGGTATCAACACACCATGAAATGATCAAGTGGTGTGTTAAG 1948
539 IleThrGlyHisPheThrAlaCysGlnValAlaGlnArgLysIleGlnGl 555
|||||
1949 ATAAGGGTCTACTTCTATGCTAGCCGCTTGCAAAAGGAAAAATTCAGGA 1998
555 uIleLeuThrGlnValLys.....GlnHisGlnGlnLysAlaLeuG 570
|||||
1999 AATACTGGCTCAGGTAAGAGACAGCAGCAGCAACAGCAGAGACAGCGC 2048
570 InSerGlyProGlnSerArgLys 579
|||||
2049 AAAGTGGCAACCCCAACCAAGAGAAAA 2077

LOCUS AF042353 3147 bp mRNA VRT 22-JAN-1998
 DEFINITION xenopus laevis KH domain-containing transcription factor B3 mRNA,
 complete cds.
 ACCESSION AF042353
 VERSION AF042353.1 GI:2801765
 KEYWORDS African clawed frog.
 SOURCE Xenopus laevis
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 3147)
 Pfaff, S.L. and Taylor, W.L.
 AUTHORS Characterization of a Xenopus oocyte factor that binds to a
 TITLE developmentally regulated cis-element in the TFIID gene
 JOURNAL Dev. Biol. 151 (1), 306-316 (1992)
 REFERENCE 92249652
 2 (bases 1 to 3147)
 Griffin, D. and Taylor, W.L.
 AUTHORS Direct Submission
 TITLE Submitted (11-JAN-1998) Biochemistry, University of Tennessee,
 JOURNAL Memphis, 858 Madison Ave., Memphis, TN 38163, USA
 FEATURES
 source 1..3147
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /cell_type="immature oocyte"
 CDS 972..2753
 /codon_start=1
 /product="KH domain-containing transcription factor B3"
 /protein_id="AAB97457.1"
 /db_xref="GI:2801766"
 /translation="MNKLYIGNLSENVSPDLESIFKESKIPFTGQFLVKSGYAFVDC
 PDEWMAKIDITLSGKVELHGKVIIEVHSVPKRSRKLQIRNPPLHLOWELDLLA
 PQSVCEQVNTDSEAVNVYTYANKEHARQLEKNGQLENYSLKVTYIPDEMAT
 IIGKEGATIRNKTQSKIDIHRENAAEKDITTHSTPEGCSACKIIMEIMQKE
 AQDTFTFEIPLKLAHNFVRLIGREGRNLKIEQDTDTKITISPLQDITLYNPER
 TITVKGSTIECAKAEVEMKKIRESYENDIAAMNLAHLIPLGLNLAALGLPPPSGGM
 PPSGASVSPPTTSASYPFGQOPESETVHLFIPALVGAIIIGKOGQHKOLSRPAGAS
 IKIAPAGDPAKLRWLITGPPEAQKAOGRHYCKLKEENFEFGPKESVKLEAHLKVP
 YAAGRVGKGKTVNELNLNTSAEVVPRDTPDENQVVKITGHFYASQLAQRKIQ
 EILQAVRQQQQQKQKQAGSQPQPRK"
 1607..1723
 /note="encodes KH domain 1"
 misc_feature 1847..2002
 /note="encodes KH domain 2"
 misc_feature 2243..2389
 /note="encodes KH domain 3"
 misc_feature 2489..2641
 /note="encodes KH domain 4"
 BASE COUNT 957 a 690 c 676 g 824 t
 ORIGIN
 alignment_scores:
 Quality: 2483.00 Length: 593
 Ratio: 4.515 Gaps: 6
 Percent Similarity: 92.749 Percent Identity: 82.125
 alignment_block:
 US-09-685-696-176 x AF042353 ..
 Align seg 1/1 to: AF042353 from: 1 to: 3147
 1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
 |||||
 972 ATGAACAAAGCTGTATATTGGAACCTTAAGCGAAAGCTCAGCCGCCCGGA 1021
 17 pLeuGluSerIlePheIleAspAlaLysIleProValSerGlyProPheL 34
 |||||
 1022 TTGTGAAAGTCTCTCAAGAGTCGAAGATCCCTTTTCACCGGCGAGTTTC 1071

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
 |||||
 1072 TGGTCAAAAGTGGATACGCGTTTGTGGACTGTCCGGATGAGACGTGGGCT 1121
 51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
 :|||
 1122 ATGAAGGCCATCGACACCCCTCTCAGGAAAGTGGAGTTGCATGGCAAGT 1171
 67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeu 84
 |||||
 1172 GATAGAGTTGACACTCAGTACCCAAAGCAAGAGTTCGAAAGCTTC 1221
 84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
 |||||
 1222 AGATAAGGAACATACCACCTCATCTACAGTGGGAGGTACTGGACAGCCTA 1271
 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
 |||||
 1272 TTAGCACAGTATGGGACAGTGGAAACTGTGACCAAGTTAACACTGATTC 1321
 117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
 |||||
 1322 AGAACTGCGTAGTAGTGAATGTAACATATGCCAATAGGACGACGCTAGAC 1371
 134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
 |||||
 1372 AAGGACTCGAAAAATTAACGCGCTATCAGCTTGAGAACTATAGCCTGAAA 1421
 151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
 |||||
 1422 GTGACGTATATACCTGATGAATGGCAACACCCGCAATCACCATCCCGCA 1471
 167 n.....ProArgGlyArgAtgGlyLeuGlyG 176
 |||||
 1472 GCTTCAGACGCCCGCAGCAGCAACACCTCAAGGAGCGGTGGGTGGGC 1521
 176 InArgGlySerSerArgGlnGlySerProGlySerValSerLysGlnLys 192
 |||||
 1522 AAAGGGACCTCGCAGCAGCGGGTCCCTGCTGCTGCTCAAGACCTTAAG 1571
 193 Pro...CysAspLeuProLeuArgLeuValProThrGlnPheValG 208
 |||||
 1572 CCTCAGTCAGAGGTTCACATGAGGATGCTGCTGCCACACAGTTTGTGG 1621
 208 YAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnT 225
 |||||
 1622 GCGCATCATTTGGAAGAGAGGTGCCACCATCAGGAACATCATAACAGAGA 1671
 225 hrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGlu 241
 |||||
 1672 CCCAGTCAAAAATAGATATTTCACCGTAAAGAAAATGCTGCTGCTGACAG 1721
 242 LysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLy 258
 |||||
 1722 AAACCGATTACGATTCACTCAACCCCTGGAAGGCTGTTTCAGTGCATGCAA 1771
 258 sSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheThrG 275
 |||||
 1772 GATCATTTAGGATCATTCGCAAGGAGGAGTCCAGTACCAAGTTCACTG 1821
 275 IuGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeu 291
 |||||
 1822 ACAAATCCCTTTAAAAATATTAGCACATAACAATTTTGTGGCCGCTTT 1871
 292 IleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspTh 308
 |||||
 1872 ATTGGGAAGAGGAAGAACCTTAAGAAAAATAGAGCAAGACACAGATAC 1921
 308 rLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluA 325
 |||||
 1922 CAAATACAAATATCTCCACTAGGAGGAGTTCAGTACACTGTAACATCCTGAAC 1971
 325 rgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341

```
|||||
1972 GGACATTACAGTAAAGGACCATTTGAGACATGTGCCAAGCCAGGAA 2021
|||||
342 GluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSerMe 358
|||||
2022 GAAGTCATCAAAAAAATTTAGGAGTCCTATGAAATGATATTGCTGCTAT 2071
|||||
358 tAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyL 375
|||||
2072 GAATCTAGAGGCACATTTGATTCCTGGTTTAAACCTGAATGCATTTGGCC 2121
|||||
375 euPheProPro...ThrSerGlyMetProProThrSerGlyProPro 390
|||||
2122 TTTTCCCAACCTCTCTGTCAGAGATGCCACCACTCTGCTGGAGTTCC 2171
|||||
391 SerAlaMetThr...ProProTyrProGlnPhe...GluGlnSerGluTh 405
|||||
2172 TCTCCGACAACATCTGCTTCTTATCCACCATTTGGGCAGCAGCCAGATC 2221
|||||
405 rGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleIleG 422
|||||
2222 AGAGACTCTTCACTCTTCACTCCAGCTTTAGCTGTGGGACCAATTATGT 2271
|||||
422 lYlYsGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSer 438
|||||
2272 GAAGCAGGACACACATCAACCAACATGTCGCGCTTTTGCAGGACATCT 2321
|||||
439 IleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetValIl 455
|||||
2322 ATTAAGATTGCACCGCAGAGGGACCTGATGCCAACCTAAGATGGTAT 2371
|||||
455 eIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrG 472
|||||
2372 CATCACTGGACCTCTCTGAAGCCCAATTTAAGGCTCAAGGAAGGATCTATG 2421
|||||
472 lYlYsIleLysGluGluAsnPheValSerProLysGluGluValLysLeu 488
|||||
2422 GTAACCTTAAAGAGAAACCTTCTTTGGACCTAAGAAAGTGAACCTT 2471
|||||
489 GluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLy 505
|||||
2472 GAGGCCACATTAAGTCGCTCATATGCTGCTGACGTGTTATTGGCAA 2521
|||||
505 sGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluValV 522
|||||
2522 AGGAGGCAAAACAGATAATGAATCAAGATTTTAAACAAGTGCAGAAGTTG 2571
|||||
522 alValProArgAspGlnThrProAspGluAsnAspGlnValValValLys 538
|||||
2572 TTGTGCCCCGTGATCAACACCATGATGAAATGATCAAGTGGTTGTTAAG 2621
|||||
539 IleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGl 555
|||||
2622 ATAACGGGTCACTTCTATGCTAGCCAGCTGTCACAAAAGGAAATTCAGGA 2671
|||||
555 uIleLeuThrGlnValLys.....GlnHisGlnGlnGlnLysAlaLeuG 570
|||||
2672 AATACTGGCTCAGTAAAGACAGACAGCAGCAGCAACAGCAGCAAGACGCG 2721
|||||
570 lNserGlyProProGlnSerArgArgLys 579
|||||
2722 AAAGTGCAACACCCCAACCAAGAAAGAAA 2750
|||||
```

seq_name: gb_ov:AF055923

seq_documentation_block: 2161 bp mRNA VRT 18-MAY-1998
LOCUS AF055923
DEFINITION Xenopus laevis RNA binding protein vera mRNA, complete cds.
ACCESSION AF055923
VERSION AF055923.1 GI:3136157
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae; Xenopodinae; Xenopus.
1 (bases 1 to 2161)
Deshler,J.O., Highett,M.I., Abramson,T. and Schnapp,B.J.
A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates
Curr. Biol. 8 (9), 489-496 (1998)
JOURNAL MEDLINE 98228351
REFERENCE 2 (bases 1 to 2161)
Deshler,J.O., Highett,M.I., Abramson,T. and Schnapp,B.J.
Direct Submission
TITLE Submitted (26-MAR-1998) Cell Biology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115, USA
JOURNAL
FEATURES
source
1. .2161
/organism="Xenopus laevis"
/db_xref="taxon:8355"
135. .1919
/codon_start=1
/product="RNA binding protein Vera"
/protein_id="AAC41285.1"
/db_xref="GI:3136158"
/translation="MNKLYIGNLSENVSPDLESLEKESKIPETGQFLVKSGVAFVDC
PDETMAKKAIDTSLGKVELHGVIEVHSVPKQRKRKLOIRNIPHLQWEVLDSLLA
QYGTENCQVNTSETAVVNVYANKHAROGLEKNGYQLENSLKYITYPIDEMAT
POAPSOLOQOQOHPQGRGFGQPARQPGGAARPKPQTEVPLRMLVPTQFVG
AIIQKEGATIRNITKQSKIDIRKENAGAAEKPIITIHSTPEGCSAAKCIWEINQK
EAODTKFTEIPLKILAHNFVRLGKESRNLKKIEODTDKTIITSPLODLTYNPE
RITYVKGSTIEPCAKAEIEIMKKIRESYENDIAMNLOAHLIPGLINLALGLFPSSSG
MPPSPVSPSTSTSPFPQOPESETVHLFIPALAVGAIGIKQGOHKKQSRFAGA
SIKIAPAEQDAKLRNVIITGPPEAQFKAGRIYKLGKEENFEKPEEKLETHIKVP
SYAAGRIGVGKGTVMELQNTSAEYVVPDQTPDENDEVVKITGHFVASQLAQRKI
QEILAQVRROOQOQKTVQSGOPQPRK"
BASE COUNT 710 a 477 c 491 g 483 t
ORIGIN
alignment_scores:
Quality: 2461.50 Length: 594
Ratio: 4.475 Gaps: 6
Percent Similarity: 92.593 Percent Identity: 81.313
alignment_block:
US-09-685-696-176 x AF055923 ..
Align seg 1/1 to: AF055923 from: 1 to: 2161
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
135 ATGAACAAGCTGTATATTGGAACCTCAGCGAAAACGCTGAGCCCAAGGA 184
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
185 TTTGGAAGTCTCTCAAGGAGTCTAAGATCCCTTTCACCGGGCAGTTTC 234
|||||
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
235 TGGTCAAAAGTGGATACGCGTTTCTGGACTGTCCTCCGGATGACAGCTGGCT 284
|||||
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
285 ATGAAGGCCATCGACACCTCTCAGGAAAAGTGGAGTTGTCATGCCAAAGT 334
|||||
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
335 GATTCAAGTCGAACATCTCGGTACCCAAAAGGCAAGGAGTGCAGAAAGCTTC 384
|||||
84 lNleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
385 AGATACGAAACATACCGCCTCATCTACAGTGGGAGGTACTGCACAGCTG 434
|||||
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||


```
source 1. .2405
/db_xref="Xenopus laevis"
/dev_stage="oocyte"
379. .2163
/notes="Vg1-RBP; Vg1 localization element binding protein"
/codon_start=1
/product="Vg1 RNA binding protein variant D"
/protein_id="AAC18598.1"
/db_xref="GI:3172449"
/translation="MNKLVIGLSENVSPDLESIFKESKIPFTGQFLVKSGVAFVDC
PDETWAKAIDITLCKVELHGKVEHSVPKRQRKLIQIRNIPHLQWEYDLSLLA
QVGTENCQVNTSTAVNVYANKHARQGLKINGIQLENTSLKVTYIPDENAT
POAPSOLOQOQQHQPOGRGFGPARQSPGAAARPKQTEVPLRMLYPTQVVG
AIIQREGATIRNITQKQIDITHKENAGAAEKPTIHSPTGGSAACKIIMEINQK
EAQDTKEIEPLKLAHNNFVRLSIEGKRNKKTEQDQTKITISPLQDLTLVNPQ
RTITVKSGETECPAKAEETMKKIRSEYENDIAAMNLOAHLIPCLNLAAGLPPSSSG
MPPSVGVSPSPSSSYPPFGQOPESETVHLFIPALAVGAILGKQOQHILQSRFAGA
SKIAPAEGDKARKRWIITGPPEAGQKQAGRIYKGLKEENFFGPKKEEVKLETHIKVP
SYAAGRVIGKGTQVNEQLNLTSAEYVVVPRDQTPDENDEVVVKITGHFVLSQLAQRKI
OEILAQVRQOQOQKTVSQGPQPRK"
BASE COUNT 763 a 537 c 537 g 568 t
ORIGIN
```

```
alignment_scores:
  Quality: 2461.50      Length: 594
  Ratio: 4.475         Gaps: 6
  Percent Similarity: 92.593      Percent Identity: 81.313

alignment_block:
US-09-685-696-176 x AF064634 ..
Align seg 1/1 to: AF064634 from: 1 to: 2405

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
379 ATGAACAAAGCTGTATTGGAAACCTCAGCGAAACGTTGAGCCCAACGGA 428
|||||

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
429 TTTGGAAAGTCTCTTCAAGGAGTCTAAGATCCCTTTCACCGGGCAGTTTC 478
|||||

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
479 TGGTCAAAAGTGGATACGCGTTTGTGGACTGTCCGGATGACACGTGGGCT 528
|||||

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
529 ATGAAGGCCATCGACACCCTCTCAGGGAAGTGGAGTTGCATGGCAAGT 578
|||||

67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
579 GATTGAAGTCGAACACTCGGTACCCAAAGGCAAGGAGTCCAAAGCTTC 628
|||||

84 InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
629 AGATACGAACACATACCGCTCATCTACAGTGGGAGGTACTGGACAGCGTG 678
|||||

101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
679 TTGGCAGATGGCAGTGGGAACCTGTGAGCAAGTGAACACTGAATC 728
|||||

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
729 AGAAACTGCAGTAGTGAATGTAACATATATGCCAATAAGGAGCACCTAGAC 778
|||||

134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
779 AAGGACTCGAAAAATTAAACGGCTTACAGCTGGAAAAATATAGCGCTGAAA 828
|||||

151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
|||||
```

```
829 GTGACGTATATACCTGATGAATGCCAACACACCGCAAGCACCATCCCAACA 878
167 n.....ProArgGlyArgArgGlyLeuG 175
|||||
879 GCTTCAGCAGCAGCGCAGCAGCAGCAACACCTCAAGGGAGCGTGGGTTTG 928
|||||

175 lyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysGln 191
|||||
929 GGCAAAAGGGGACCTGCCAGGAGGGATCCCTGCTGCTGCTGCAAGACCT 978
|||||

192 LysPro...CysAspLeuProLeuArgLeuLeuValProThrGlnPheVa 207
|||||
979 AAGCTTCAGACAGAGTTCCTGCTGAGAATGCTGGTTCACACACAGTTGT 1028
|||||

207 lGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
|||||
1029 AGGGCGCATCATTTGGGAAAGAGGTGCTTACCATCAGGAACATCACTAAC 1078
|||||

224 InThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAla 240
|||||
1079 AGACCCAGTCAAAAATAGATATTACCCGTAAGAAAATGCAGGTGCTGCA 1128
|||||

241 GluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCy 257
|||||
1129 GAGAAACCGATTACAATTCACTCAACCCCTGAAGGCTGCTCAGCTGCATG 1178
|||||

257 sLysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheT 274
|||||
1179 TAAGATCATTTATGGAGATCATGCAGAGGAAGCTCAGGATACCAAGTTCA 1228
|||||

274 hrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArg 290
|||||
1229 CTGAAGAAATCCCTTAAAAATATTAGCACATAACAACTTTGTTGGCGGT 1278
|||||

291 LeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAs 307
|||||
1279 CTTATTTGGTAAAGAGGAGGAAGCACTTAAAGAAAATTGAGCAAGATACAGA 1328
|||||

307 pThrLysIleThrIleSerProLeuGlnLeuThrLeuTyrAsnProG 324
|||||
1329 TACAAAATCACAATATCTCCACTACAGGACTTGACACTGTACAAATCCAG 1378
|||||

324 luArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaGlu 340
|||||
1379 AACGGACCATTTACAGTAAAGGAGCAGCATTTAGCCATGTGCCAAAGCCGAG 1428
|||||

341 GluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaSe 357
|||||
1429 GAAGAAATCATGAAAAAATTTAGGAGTCCCTATGAAAATGATATTGCTGC 1478
|||||

357 rMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuG 374
|||||
1479 TATGAATCTCCAGGCACATTTGATTCTCTGTTTGAACCTGAATGCATGG 1528
|||||

374 lyLeuPhePro...ProThrSerGlyMetProProThrProThrSerGlyPro 389
|||||
1529 GCCTTTTCCCATCATCTCATCAGGAATGCCACCACCTTCTGTTGGAGTT 1578
|||||

390 ProSer...AlaMetThrProThrProGlnPhe...GluGlnSerG 404
|||||
1579 CCCTACCTACCTCATCTACTCTTATCCACCATTGGGGCAGCAGCCAGA 1628
|||||

404 uThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleI 421
|||||
1629 GTCAGAGACCGTTTCATCTCTTATCCAGCTTTAGCTGTGGAGCGATTA 1678
|||||

421 leGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 437
|||||
1679 TTGAAAAGCAAGGACACACATCAAAACAACTATCACGTTTTCAGGAGCA 1728
|||||

438 SerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetVa 454
|||||
1729 TCTATTAGATTGACCTGCGGAGGACCTGATGCCAAACTAAGAAATGGT 1778
|||||
```

454 l l l e l l e t h r G l y P r o P r o G l u A l a G l n P h e L y s A l a G l n G l y A r g I l e t 471
|||||
1779 G A T C A T C A C T G G A C C T C C T G A A G C C C A A T T A A G G C T C A A G G A A G G A T C T 1828
|||||
471 y r G l y L y s l e l y s G l u G l u A s n P h e V a l S e r P r o L y s G l u G l u V a l L y s 487
|||||
1829 A T G G T A A A C T G A A G A A G A A A A C T T C T G G A C C T T A A G A A G A A G T G A A 1878
|||||
488 L e u G l u A l a H i s l e A r g V a l P r o S e r P h e A l a A l a G l y A r g V a l l e G l 504
|||||
1879 C I T G A C C C A C A T T A A G T G C G T C A T A T G T G T G G A C G G T A T T G G 1928
|||||
504 y l y s G l y G l y T h r V a l a s n G l u L e u G l n A s n L e u S e r S e r A l a G l u V 521
|||||
1929 C A A G A G G A A A A C A G T A A T G A A T T C A G A A T T T A C A A G T G C A G A G G 1978
|||||
521 a l v a l V a l P r o A r g A s p G l n T h r P r o A s p G l u A s n A s p G l n V a l V a l 537
|||||
1979 T T G T T G C C C G T G A T C A A A C A C C A G A T G A A A T G A T G A A G T G G T G T 2028
|||||
538 L y s l e t h r G l y H i s P h e T y r A c y s G l n V a l A l a G l n A r g L y s l e G l 554
|||||
2029 A G A T A A C C G G T C A C T T C T A T G C A G T C A G T T G C A A A G G A A A T T C A 2078
|||||
554 n G l u l l e L e u t h r G l n V a l L y s G l n H i s G l n G l n L y s A l a L 569
|||||
2079 G G A A A T A T T G G C T C A G T A A G A A G A C A G C A G C A C A C A G A A A C A G 2128
|||||
569 e u G l n S e r G l y P r o P r o G l n S e r A r g L y s 579
|||||
2129 T G C A A A G T G G A C A A C C C A C C G A G A G A A A A 2160
|||||

seq_name: gb_ov:AF161270

seq_documentation_block: 2011 bp mRNA VRT 27-JUL-1999
LOCUS AF161270
DEFINITION Danio rerio Vgl RNA binding protein mRNA, complete cds.
ACCESSION AF161270
VERSION AF161270.1 GI:5596631
KEYWORDS
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 2011)
AUTHORS Zhang,Q., Yaniv,K., Oberman,F., Wolke,U., Git,A., Fromer,M.,
Taylor,W., Meyer,D., Standart,N., Raz,E. and Yisraeli,J.K.
Vgl RBP intracellular distribution and evolutionarily conserved
expression suggest multiple roles during development
Mech. Dev. (1999) In press
REFERENCE 2 (bases 1 to 2011)
AUTHORS Oberman,F., Fromer,M. and Yisraeli,J.K.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1999) Anatomy and Cell Biology, Hebrew University
Hadasah Medical School, POB 12272, Jerusalem 91120, Israel
FEATURES
source 1. .2011
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="linkage group 19"
/map="between Z3023 and Z15451"
1. .1749
/note="Vgl RBP"
/codon_start=1
/product="Vgl RNA binding protein"
/protein_id="AAD45610.1"
/db_xref="GI:5596632"
/translation="MNKLYIGNVSEQAALDLESIFQWKIPFSAPFLVKSQAFVDC
PDEKVARADITLSGKVLHGKVLVEHSVPKRSCKLQIRNIPPHQWELDGLIA
QYGVSECEQVNTDTFAVNVRYRKADQAREAMDKLNGFLMENAALKVSIYPDETAA
ADAPAVGRRGRFNPGRPGPSLGRPKLQSDVPLRLLLVPTQFVGAIIIGKEGATIR

NITKQTHSKIDIHRKENAGAAEKPIVTHSTPEGCSSACRNIMEIMQKRAIDTKITEEI
PLKILAHNNFVGRLLIGRNLKKIEQDQTDKTIITSPLODITLYNPRTIIVKGTFLDA
CAKAEETIMKVRRESYENDVAAMHLQSNLIPGLNINLALGFPGAASGSPVSGPP
PQAQGYOSFGAOMESFTVHLFIPALVAGIIGKOGHIIKOLSRFAGASIKIAPADGI
DAKQMWIISGPEAOPKAQGRIFGKLKEENFGPKKEVKLEAHIKVPSEARVIGK
GGKTVNELONLTSAEVVVPRDQTPDENDQVVVKITGHFYASQLAKRQIQLIISQVRRQ
QQPKPSAAGPPVARRK"

BASE COUNT 581 a 494 c 501 g 435 t
ORIGIN

alignment_scores:
Quality: 2261.50 Length: 584
Ratio: 4.299 Gaps: 5
Percent Similarity: 90.068 Percent Identity: 77.055
alignment_block:
US-09-685-696-176 x AF161270 ..
Align seg 1/1 to: AF161270 from: 1 to: 2011

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
1 ATCAATAAGCTGTACATCGGAAACGTGACGAGCGAGCGCGCGTGA 50
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
51 CITGGAAAGTATCTTTGACGAGCTGGAATACCTTTTCAGTGCTCCCTCC 100
|||||
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
101 TCGTAAAAAGTGGCTACGGCTTTGTGGATTGCCCGCAGCAAGAGGTAGC 150
|||||
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 ATGAGGGCCATTGACACGCTTTTCAGGAAAAGTAGAACTCCACGGAAGAAGT 200
|||||
67 olleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 CITAGAAGTGGAGCAGCTCTGTCCCTAAGCGACAAGAGAGCTGTAAAGTCG 250
|||||
84 lnleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATCAGGAATATCCACCTCATCATGCTGAGTGGGAAGTTTATAGCGGTTG 300
|||||
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 TTAGCACAGTATGGCAGCTGGAAGCTGTGAACAAGTCAACACTGCACAC 350
|||||
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
351 CGNAGCAGCAGCTGCTCAATGTGGCATATGGTCCCAAGGACCAAGCCAGAG 400
|||||
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCAATGACAAATTAATATGGCTTCCTTATGGAGAATACGACACTAAAG 450
|||||
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||
451 GTTTCCTATATCCAGATGAGCTGCTGCAGCTGTATGCCCTCCTCCCTCCT 494
|||||
167 nProArgGlyArgArgGlyGlnArgGlySerSerArgGlnGlyLys 184
|||||
495 TGTGGGTGGCAGGAGAGGCTTTAAACCCACGCTGGGCCACCTCGCCAGGCT 544
|||||
184 erProGlySerValSerLysGlnLys...ProCysAspLeuProLeuArg 199
|||||
545 CTCCCAGTCTGGCGCTAGACCTAAACTACAGTCAGACGTCGCTCTACGC 594
|||||
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
|||||
595 TTGCTGGTACCCACCGAGTTGTAGGAGCTATTATTGGCAAGGAAGGTGC 644
|||||

```
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisa 233
|||||
645 TACCATCGCAACATCAACCAAGCAGACCCACTCAAGATTGACATTCATC 694
|||||
233 rgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
|||||
695 GTAAGGAGAACGACAGGACGACAGAGAACCAATCACTGTTCACTCAACA 744
|||||
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
|||||
745 CCAGAGGCTCGAGTTCGCTCGTGAACATCATGAGATCATGCAGAA 794
|||||
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
|||||
795 AGAGCAATCGACACAAAGATCACTGAAGAAATTCCTCAAAATTTGG 844
|||||
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
|||||
845 CTCATAACAATTTGTTGGCGCCTCATTTGAAAGAGGGACGCAACTTG 894
|||||
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlu 316
|||||
895 AAGAAATCGAACAAGATCTGACACCAAGATCACATCTCACCTCTGCA 944
|||||
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
|||||
945 GGATCTGACCTTGTATACCCAGAGCGCACCATTAATCTGTTAAAGGAAGCT 994
|||||
333 alGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGlu 349
|||||
995 TAGATGCTGTGCAAAAGCTCAAGAAAGAAATCATGAAGAAAGTCCGGAG 1044
|||||
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
|||||
1045 TCCTATGAGAACGATGTAGCTGCCATGCACCTTCAATCACTACCTGATCC 1094
|||||
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhePro...ProThrSerGly. 381
|||||
1095 AGGGCTCAATCTGAATGATGCTGGGTTATTTCTCTGGTGCAGCGAGTGGAG 1144
|||||
382 .....MetProProThrSerGlyProProSerAlaMetThrProPro 396
|||||
1145 GGATTTACCTCTCTGTGTATCAGACACCCGCTGGGGCTCAAGCTGGA 1194
|||||
397 TyrProGlnPhe...GluGlnSerGluThrGluThrValHisGlnPheIl 412
|||||
1195 TATCAATCATTTGGGGCCACATGAGTGCAGAACTGTGACCTCTTCAT 1244
|||||
412 eProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleL 429
|||||
1245 TCCTGCACTGGCAGTTGGTGTCTATTATCGCAAGCAGGGCCCAACACATTA 1294
|||||
429 ysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445
|||||
1295 AACAGCTCAGCGCTGCTGCTGGTGCCTCAATCAAAATTCGCCCTGCTGAT 1344
|||||
446 AlaProAspAlaLysValArgMetValIleIleThrGlyProGluAla 462
|||||
1345 GGAATTTAGTCCAGCAAGAAATGGTATTATTCTGGACACACAGAGGC 1394
|||||
462 agLnpHelLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsn 479
|||||
1395 ACAGTTTAAGGCTCAGGACCTATCTTTGGAAAGCTTAAAGAGGAGAACT 1444
|||||
479 heValSerProLysGluGluValLysLysLeuAlaHisIleArgValPro 495
|||||
1445 TCCTCGGACCCAAAGAGAGGTTAAATAGAGGCTCATATAAAGTGCCTA 1494
|||||
496 SerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnG 512
|||||
1495 TCATTCGCTGCTGCAGAGTCTATTTGGTAAGAGGACCAAAACGGTGAACGA 1544
|||||
512 uLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnThrP 529
|||||
```

```
|||||
1545 ACTACAGAACTTGACCACTGGGAGGTGGTCCACAGGGACACAGAC 1594
|||||
529 roAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyrAla 545
|||||
1595 CTGATGAAAATGACCAAGTAGTAGTAAATAATCACAGGCCACTTCTATGCA 1644
|||||
546 CysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlu 562
|||||
1645 AGCCAGCTGGCCCGCCAGAGAAATCCAGGAGATTATCTCTCAGGTGAGCGG 1694
|||||
562 nHisGlnGlnGlnLysAlaLeuGlnSerGlyProGlnSerArgArgL 579
|||||
1695 GCAACAGCAGCCCAAGCCCTCAGCGCAGGTCCACCACTCGCCAGGAGGA 1744
|||||
579 ys 579
||
1745 AA 1746

seq_name: gb_ov:AF026527
seq_documentation_block:
LOCUS AF026527 2021 bp mRNA VRT 30-OCT-1997
DEFINITION Gallus gallus zipcode-binding protein (ZBP1) mRNA, complete cds.
ACCESSION AF026527
VERSION AF026527.1 GI:2570920
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2021)
AUTHORS Ross, A.F., Oleynikov, Y.S., Kislauskis, E.H., Taneja, K.L. and
Singer, R.H.
TITLE Characterization of a beta-actin mRNA zipcode-binding protein
JOURNAL Mol. Cell. Biol. 17 (4), 2158-2165 (1997)
MEDLINE 9722007
REFERENCE 2 (bases 1 to 2021)
AUTHORS Oleynikov, Y.S., Ross, A.F. and Singer, R.H.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
source
1..2021
/organism="Gallus gallus"
/db_xref="taxon:9031"
/cell_type="fibroblast"
/dev_stage="embryo"
1..2021
/gene="ZBP1"
129..1859
/gene="ZBP1"
/function="binds a zipcode element in beta-actin mRNA
3'UTR and is involved in its localization"
/notes="beta-actin mRNA zipcode-binding protein"
/codon_start=1
/product="zipcode-binding protein"
/protein_id="AAB82295.1"
/db_xref="GI:2570921"
/translation="MNKLYIGNLNSVTPADLEKVFNDHKISFGQFLVKSGYAFVDC
PDEQWAMKAIEFTSGKVELHGOLEIHSVPKRSRKIQIRNIPQLRWEVLDGLLA
OYGTENCBOVNTDSETAVNVYTNREOTROAIMKLNHOLENHVLYKVSYPDEQSV
QGPENGRGCGFARGAPROGSPVTAGAPVKQPDPLRLVPTOVVGALIGEGATI
RNITKOTQSKIDVHRKENAGAAEKASISHTSTEGCSACKNMILEIMQKEAKDKTDAE
VPLKILAHNNFVGRLLIGERNLKRVEQDETAKTITISSLODLTYINPERTITVKGSIE
NCCKAQOEIMKKVREAYENDVAAMSLQSHLIPGLNLAAYGLFPASSNAVPPPPSSVSG
AAPYSFMPPEQETVHFVIPAQAVGAIIGKQGHQIKQLSRFASASIKIAPPETPSKV
RMVLTGPPEAQFKAQGRIVGKLEENFFGPEEVLKETHIRVPASAAQGRVIGKQKGT
VNELQNLTAAEVVPDQTPDENEQVIVLIGHIFYASQMAQRKIRDILIAQVKQHQKG
QSGQLQARRK"
138..155
/misc_feature
/gene="ZBP1"
```

```

misc_feature      /note="encodes RNP2 RNA binding domain"
236..260
/gene="ZBP1"
misc_feature      /note="encodes RNP1 RNA binding domain"
354..374
/gene="ZBP1"
misc_feature      /note="encodes nuclear localization signal; NLS"
621..632
/gene="ZBP1"
misc_feature      /note="encodes nuclear localization signal; NLS"
747..773
/gene="ZBP1"
misc_feature      /note="encodes KH-domain #1"
990..1016
/gene="ZBP1"
misc_feature      /note="encodes KH-domain #2"
1062..1088
/gene="ZBP1"
misc_feature      /note="encodes REV-like nuclear export signal; NES"
1374..1400
/gene="ZBP1"
misc_feature      /note="encodes KH-domain #3"
1620..1646
/gene="ZBP1"
misc_feature      /note="encodes KH-domain #4"
525 a 562 c 583 g 351 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 2219.50      Length: 584
  Ratio: 4.196          Gaps: 7
  Percent Similarity: 90.582  Percent Identity: 74.829

alignment_block:
US-09-685-696-176 x AF026527 ..
Align seg 1/1 to: AF026527 from: 1 to: 2021

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
129 ATGAACAAGCTGTACATCGGACCTTAAACGAGAGCGTGACTCGGCCGA 178
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
179 CTTGGAGAAAGTTTCAACGACCAAGATCTCCTTCAGCGGCGACTTCC 228
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
229 TGGTCAAGTCCGGCTATGCTTTTGTGGACTGCCCGCAGCAGTGGGCC 278
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
279 ATGAAGCCATCGAAATCTTTTCGGGGAAGTGGAGCTGCATGGAACA 328
67 orLeuGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
329 GCTTGAGATTGAACACTCAGTGCCTTAAAGCAAGAGCGCGAAATTC 378
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
379 AAATCCGAATATCCACCCGCTGCGATGGAGGTTCTGGATGGCTTG 428
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
429 CTAGCAACAATATGGCAATCTGAGAAAATCTGGACAAAGTGAACACAGAC 478
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
479 TGAGACCGCTGTGGTTAATGTCACCTACACAACCGGGAGCAGCAGGC 528
134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

529 AAGCCATCATGAAGTTAAACGGCGCACCGCTGGAGAACACACGCTGCTGAAA 578
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
579 GTCTCTACATCCTGATGAGCAGTCCGTG...CAGGGCCAGAGATGG 625
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
626 CGGGCGGGGT...GGCTTTGGGGCCCGAGGTGCCCCCGCAGGGCT 669
184 exPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
670 CCCCTGTCACTGCAGGGGCTCCAGTCAAGCAGCAGCCCGTGGACATCCG 719
198 LeuArgLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
720 CTCCTGCTGCTGGTGGCCAGCAGTATGTTGGTGCCATCATTTGGCAAGGA 769
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
770 AGGGGCCACCATCAGGAACATTACCAAGCAGACACAGTCCCAAGATTGATG 819
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
820 TGCACCGTAAAGGAATGCAGGAGCTGCAGAAAAGCTATCAGCATCCAC 869
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
870 TCACCCCGGAGGGCTGCTGCTGCTGCAAAATGATCTTGGAGATCAT 919
264 tHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
920 GCAGAAGGAGGCAAGACACAGACAGTGTGATGAAGTGCCTCTGAAA 969
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
970 TCTTGGCCCATAAACATTTTGTGGGGCGCTGATTGGCAAGAAGGCGG 1019
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1020 AACTTGAAGAAAGTGGAGCAGGATACTGAGACAAAATACCATCTCATC 1069
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1070 CTTGAGGACCTGACCTGTACAAACCCGAAAGAACGATCACAGTGAAG 1119
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1120 GCTCTATCGAGAACTGCTGCAAGCAGCAGCAGGAGATCATGAAGAAGTG 1169
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1170 AGAAGAGCTTACGAGAACGATGTTGAGCCATGAGCCTGCAATCTCATCT 1219
364 ulleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerG 381
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1220 CATCCCTGGCTCAACTGCTGCTGAGTGGCCCTTCCCTGCTCCCTCCA 1269
381 lyMetProProProThrSerGlyProProSerAlaMetThr.....Pro 395
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1270 ATGACAGTACCTCTCT.....CTCTCCAGCAGCAGCTCTCCGGGGTGTG 1310
396 ProTyrProGlnPheGluGlnSerGluThrGluThrValHisGlnPheI 412
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1311 CCATAGCTCTCTTCATGCTCCGGAGCAGGAGCCGTCACAGCTCTTCAT 1360
412 eProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIle 429
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1361 CCCTGCCCAAGCGGTTGGTCCCATCATTTGGCAAGAGGGCCAGCACATCA 1410
429 yGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445
|||||:  |||||||  |||||||  |||||||  |||||||  |||||||
1411 AGCAGCTCTCCCGGTTCCGAAGTGCCTCTATTAAGATTGCACCCCGGAG 1460

```



```

|||||
381 AGATCCGCAATATTCACCTCCAGCTCCGATCGGAAGTGTAGATAGCCTG 430
|||||
101 LeuValGlnThrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
431 CTGGCTCAGTACGAGTACGAGTGAAGTGTGAGCAAGTGAACACTGAAAG 480
|||||
117 rGluThrAlaValAlaValAsnValThrThrSerSerLysAspGlnAlaArg 134
|||||
481 TGAGACAGCGGTGGTCAACGTCACTACTAACCAGGAGCAGCAGCAGGC 530
|||||
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
531 AAGCTATCATGAAGTAAATGGCCATCACTGAGAGAACCATGCGCTGAAG 580
|||||
151 ValAlaThrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
581 GTCTCTACATCACTGATGAGCAGATAACA...CAAGGTCTCTGAGAAATG 627
|||||
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly 184
|||||
628 GCCTCGTGA...GGCTTTGGTCTCGGGCCAGCCCGGCAAGGTT 671
|||||
184 erPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
|||||
672 CGCCCTGGCAGCAGGCGGTCCAGCCAAAGCAGCAGCAGTGGACATCCCT 721
|||||
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
|||||
722 CTCGGCTCTGTGTGCTACGCAAGTATGAGCGCTATCATTTGGCAAGGA 771
|||||
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
|||||
772 GGGTGCACCATCCGAACATCACAAACAGCAGCAGTCCANATAGACG 821
|||||
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
|||||
822 TGCATAGGAAGGAATGGGGCGCTCGCGAGAAGGCCATCAGCGTGCAT 871
|||||
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
|||||
872 TCAACCCCTGAAGCTGCTCTCCGCGTCAAGATGATCTTTGGAGATTAT 921
|||||
264 tHisLysGluAlaGlnAspLysPheThrGluGluIleProLeuLysI 281
|||||
922 GCACAGGAGGCAAGSACACCAACAGCGCAGATGAAGTTCCCTCGAGA 971
|||||
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
|||||
972 TCCTGGCTCATAACTTCTGTCGGCGACTCATTTGGCAAGGAAGCGCG 1021
|||||
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||
1022 AACCTGAAGAGGTGGAGCAGGACAGACAGACCAAGATCACCATCTCATC 1071
|||||
314 oLeuGlnGluLeuThrLeuThrAsnProGluArgThrIleThrValLysG 331
|||||
1072 GCTCCAGGACCTCACGCTCTATAACCTCGAGAGACCACATCTGTGAAG 1121
|||||
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||
1122 GCGCCATTGAGAACTGTTCAGGGCGCAGCAGGAGATCATGAAGAAGTT 1171
|||||
348 ArgGluSerTyrcGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
|||||
1172 CGAGAGGCTTACGAGAACGACGTCGGCGCCATGAGCTTCAGTCCACCT 1221
|||||
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
|||||
1222 CATCCCTGGCTTAACCTGGCTGCTAGTCTCTCCAGCTTCATCCA 1271
|||||
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
:: |||||

```

```

1272 GCCTGTCTCCCTCT.....CTCCCCAGCAGTGTCTACTGGGGCTGCT 1312
|||||
396 ProTyrrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
|||||
1313 CCTATAGCTCTCTTCATGCAGGCTCCGAGCAGAGATGGTACAAAGTGT 1362
|||||
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisI 428
|||||
1363 CATCCCCGCCAGGCTGTGGGCCCATCATTTGGCAAGAGGCCAGCACA 1412
|||||
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
|||||
1413 TCAACAACACTCTCCGCTTCGCGAGCGCTCCATCAAGATTGCTCCACA 1462
|||||
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
|||||
1463 GAACACCTGACTCCAAAGTTCGAATGCTGCTATCATCTGACCCCCAGA 1512
|||||
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrrGlyLysIleLysGluAla 478
|||||
1513 GGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGCAAACTAAAGAAGAGA 1562
|||||
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
|||||
1563 ATTTCCTTTGGTCCCAAGGAGAAAGTAAAGCTAGAGACCCACATACGG 1612
|||||
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValas 511
|||||
1613 CCGCTTCAGCAGCCGCGCTCATCGGCAAGGCGCAAAACGGTGAA 1662
|||||
511 nGluLeuGlnAsnLeuSerSerAlaGluValValValProArgAspGlnT 528
|||||
1663 TGAGCTGCAGAACTTGACTGCAGCTGAGGTGGTAGTGCCAAAGAGACC 1712
|||||
528 hrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTy 544
|||||
1713 CCCGGATGAGAAACGACCAAGTCAATTTAAGATCATCGGACATTTTAT 1762
|||||
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLy 561
|||||
1763 GCAGCGCAGATGGCTCAGCGGAAGATCCGAGACATCTCTGCTCAAGTAA 1812
|||||
561 sGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
|||||
1813 GCAA...CAGCACCAGAAGGA...CAGACCAACTGGCCCGGCTCGGA 1856
|||||
578 rGlys 579
|||||
1857 GGAAG 1861

seq_name: gb_pr:AF117106
seq_documentation_block: 2130 bp mRNA PRI 26-JAN-1999
LOCUS AF117106 Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA, complete
DEFINITION cds.
ACCESSION AF117106
VERSION AF117106.1 GI:4191607
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2130)
AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.
TITLE A family of insulin-like growth factor II mRNA-binding proteins
represses translation in late development
JOURNAL Mol. Cell. Biol. 19 (2), 1262-1270 (1999)
MEDLINE 99108099
REFERENCE 2 (bases 1 to 2130)
AUTHORS Nielsen,J., Christensen,J., Lykke-Andersen,J., Johnsen,A.H.,
Wewer,U.M. and Nielsen,F.C.

```



```
|||||
1292 TCAACAGCTCCCGGTTTGGCAGCGCTCCATCAAGATTCACACACC 1341
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
|||||
1342 GAACACCTGACTCCAAAGTTCGTATGTTATCATCTACCTGGACCGCAGA 1391
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluL 478
|||||
1392 GGCCCAATTCAGGCTCAGGAAGAATCTATGCAAACTCAAGGAGGAGA 1441
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
|||||
1442 ACTTCTTGGTCCCAAGGAGGAGTGAAGCTGGAGACCCACATACCTGTG 1491
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAs 511
|||||
1492 CCAGCATCAGCAGCTGGCGGGTTCATGGCAAGGTGGAAAACGGTGAA 1541
511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
|||||
1542 CGAGTTGCAGAAATTCACGCGCAGCTCAGGTGGTAGTACCAAGAGACCAGA 1591
528 hrProAspGluAsnAspGlnValValLysIleThrGlyHisPheTyr 544
|||||
1592 CCCTCATGAGAACGACGACGAGTCTCGTGAATCATCGCAGCATTTCTAT 1641
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVal 561
|||||
1642 GCAGTCAGATGCTCAACGGAAGATCCGAGACATCTCGCCAGGTAA 1691
561 sGlnHisGlnGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArg 578
|||||
1692 GCAG...CAGCATCAGAAGGA...CAGAGTACACGACCCGACGACGGA 1735
578 rGlys 579
|||||
1736 GGAAG 1740
```

seq_name: gb_pr:AF198254

```
seq_documentation_block: 2381 bp mRNA PRI 02-MAR-2000
LOCUS AF198254 Homo sapiens mRNA-binding protein CRDBP mRNA, complete cds.
DEFINITION AF198254
ACCESSION AF198254
VERSION AF198254.1 GI:7141071
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2381)
Ioannidis,P., Trangas,T., Dimitriadis,E., Samiotaki,M.,
Panoutsakopoulos,G., Kyriazoglou,I., Voutzoulas,S.,
Tsilapalis,C.M., Kittas,C., Agnantis,N. and Pandis,N.
Ectopic expression of a KH-domain containing protein, highly
homologous to both human IMP-1 and mouse CRD-BP, in benign and
malignant mesenchymal tumors
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2381)
AUTHORS Ioannidis,P., Trangas,T., Dimitriadis,E. and Samiotaki,M.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) G. Papanikolaou Research Center, St Savvas
Hospital, 171 Alexandras Avenue, Athens, Attica 11522, Greece
FEATURES
source
1. .2381
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21.3-q22"
/cell_line="MCF-7"
306. .2039
/feature="KH-domain containing protein"
```

```
/codon_start=1
/product="mRNA-binding protein CRDBP"
/protein_id="AAF37203.1"
/db_xref="GI:7141072"
/translaton="MNKLYIGNLNEVTPADLEKVFABHKISYSCOFVKSQYAFVDC
PDERWAMKAIEITFSGKVELQGRLEIEHSYFKKORSRKIQIRNPPQLRWELVDSLLA
QYGTVENCEQVNTSEAVVNTYSNREQTQAOIMKLNHOLENHALKVSYPDEQIA
QPGNRRGGFGSGRQSPVAAQAPAKQOQVDIPURLLVPTQYVGAIIIGKEGATI
RNIKTQTSKIDVHRKENAGAERAISVHSTPEGSSACKMILETHMKKADTKTADE
VPLKTLAHNFVRLIGKEGRNLKKEODTETKITISSLODLTLNPERTITVKAIE
NCCRAEQEIMKKVREAVENDVAAMSLOSHTPGLNLAAGLFPASSASVAPPPSSVTG
AAPYSSPQAPQEQMVQVFIQAQAVGAILIKKKQHILKHSFASASIKIAPPEPDSK
VRMWIITGPPEAQKQAGRIYGLKLEENFFGPKKEVKLETHIRVPASAAGRVIGGGK
TVNELQNLTAAEVVPVPRDQTPDENDQIVIKIIGHFYASQAQRKIRDILAQVKQHQK
GQSNAQAARRK"
```

BASE COUNT 596 a 687 c 664 g 434 t
ORIGIN

alignment_scores:

Quality: 2181.00 Length: 585
Ratio: 4.154 Gaps: 8
Percent Similarity: 89.744 Percent Identity: 74.017

alignment_block:

US-09-685-696-176 x AF198254 ..

Align seg 1/1 to: AF198254 from: 1 to: 2381

```
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
306 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGGCTGACCCCGCGGA 355
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
356 CTTGGAGAAAGTGTTCGGGAGCACAAAGATCTCTACAGCGCGCAGTTCT 405
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
406 TGGTCAATCCGGCTACGCTTCGTGGACTCCCGGACGAGCAGCTGGCG 455
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
456 ATGAGGCCATCGAAATCTTCCTCCGGGAAAGTAGAATTACAAGGAAAAAC 505
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
506 CTTAGAGATTGAACATTCGGTCCCAAAAAAAGAGGAGCGCGAAATTC 555
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
556 AAATCCGAATATTTCCACCCAGCTCCGATGGGAAGTACTGGACAGCTG 605
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
606 CTGGCTCAGTATGTCAGTAGAGAACTGTGAGCAAGTGAACACCGAGAG 655
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
656 TGAGAGCGGAGTGGTGAATGTCACCTATTCCACCGGGAGCAGACACGAG 705
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
706 AACCCATCATGAGCTGAATGGCCACCATGTTGGAGAACCATGCCCTGAAG 755
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
756 GTCTCTACATCCCGATGAGCAGATAGCA...CAGGACCTGAGATGG 802
167 nProArgGlyArgArgGlyLeuGlnArgGlySerArgGlnGlyS 184
|||||
803 GCGCCGAGGG...GGCTTTGGCTCTCGGGGTCAGCCCGCCGAGGCT 846
```


SPSPQRAQRGDHSSRQGHAPGCTSQARQIDFPLRLVPTQFVGAIGKEGLTKNI
TKQTSQVDRHREKNSGAEEKPVTIHATPEGTSEACRMILEIMQKEDETKLAEEIPL
KILAHNGRLRIGKEGRNLKRIEHTGKTITSSLODLSTIYNPERTITVKGTVEACA
SAETEMKKLREAFENMLAVNOQANLIPGLNLSALGIFSTGLSVLSPAPGRGAPPA
APYHPFTTHSGYFSSLYPHQGFPPPHHHSYPEQFIVNLRPTQAVGAILGKGAHLK
QLARFAGASIKIAPAEQPDVSERNVITGPEAQFKAGRIFGKLKEENFNPKEEVK
LEAHIRPSTAGVIGKGTGKTVNELQNLTSAEVIVPRDQTPDENEVIVRIIGHFFA
SOTAKRIRRETVQGVKQOQKYPQGVASQSRK*
BASE COUNT 565 a 552 c 532 g 361 t
ORIGIN

alignment_scores:
Quality: 1950.50 Length: 591
Ratio: 3.817 Gaps: 8
Percent Similarity: 86.464 Percent Identity: 65.313
alignment_block:
US-09-685-696-176 x AF117107 ..
Align seg 1/1 to: AF117107 from: 1 to: 2010

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
50 ATGAACAAGCTTTACATCGGGAACCTTGAGCCCGCGCTCACCCGCCGACGA 99
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
100 CCTCCGGCAGCTTTGGGACAGGAGCTGCCCTGGCGGACAGGTCC 149
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
150 TGTGAAGTCCGGCTACGCTTCTGCTGACTACCCCGACACCACTGGGCC 199
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
200 ATCCGCGCCATCGAGACCTCTCGGGAAGTGAATTCATGGGAAAAAT 249
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
250 CATGGGAAGTTGATCTAGTCTCTAAAAGCTAAGGACGAGCAAAATTC 299
84 InleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
300 AGATTTCGAACATCCCTCTCACCCTCAGTGGGAGGTGTGGATGCAT 349
101 LeuValGlnTyrGlyValValGluSerCysGluInValAsnThrAspSe 117
350 TTGGCTCAATATGGGACGTGGGAGATGTGGAACAAGTCAACACACAC 399
117 rGluThrAlaValAlaValAsnValThrTyrSerSerLysAspGlnAlaArg 134
400 AGNAACCGCGTGTGTCACGTACATATGCAACAGAGAAGACAAATA 449
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
450 TAGCCATGGAGAAGCTAAGCGGSCATCAGTTTCAGAACTACTCCTTCA 499
151 ValAlaTyrIleProAspGluMetAlaAlaGlnInAsnProLeuGlnG 167
500 ATTTCCTACATCCCGATGAAGAGGTGAGCTCCCTTCGCGCCCTCAG.. 547
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly. 183
548 ...CGAGCCCGACG...GGGACCACTCTTCCCGGAGCAAGGCC 587
184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
588 ACGCCCTGGGGCATTCTCAGGCAGACAGATTGATTCCCGCTGCGG 637
200 LeuLeuValProThrGlnPheValGlyAlaIleGlyLysGluGlyAl 216
638 ATCCTGGTCCCAACCCAGTTTGTGTGTCATCATCGGAAGAGGAGGCTT 687

216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 233
688 GACCAATAAGAACATCATAAGCAGACCCAGCTCCCGGTAGATATCCATA 737
233 rGlyGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
738 GAAAGAGAACTCTGGAGCTGCAGAGAAGCTGTCCACCATCCATGCCACC 787
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethHis 266
788 CCAGAGGACATTCTGAAGCATGCCGATGATTTGAAATCATGAGAA 837
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeu 283
838 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTCTGAAAATCTTGG 887
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
888 CACACAATGGCTTGGTGGAGACTGATTGGAAAAGAACGACGAAATTTG 937
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG 316
938 AGAAAATTGAACATGAACACAGGACCAAGATAACAATCTCATCTTTGCA 987
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
988 GGATTTCAGCATATACACCCGGAAGAACCATCCTGTGAAGGGCACAG 1037
333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
1038 TTGAGGCTGTGCGAGTGTGAGATAGAGATTATGAAGAAGCTGCGTGAG 1087
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1088 GCCTTTGAAAATGATGTGCTGTAAACCAACAGCAATCTGATCC 1137
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1138 AGGTTGAACCTCAGCGCATCTTTCACAGGACTGTCCGTGC 1187
377ProThrSer.....GlyMetProProThr..... 386
1188 TATCTCCACGAGCGCCCGCGGAGCTCCCGCGCTGCCCTTACCAC 1237
387SerGlyProProSerAlaMetThrPro..... 395
1238 CCCTTCACCTACCCACTCCGATATCTTCCAGCTGTACCCCATCACCA 1287
396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1288 GTTGGCCCGTTCGCCATCATCTCTTATCCAGAGCAGAGATTTGTA 1337
409 isGlnPheIleProAlaLeuSerValGlyAlaIleLeGlyLysGlnGly 425
1338 ATCTCTTCATCCCAACCCAGGCTGTGGCGCCATCATCGGGAAGAAGGG 1387
426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1388 GCACACATCAACAGCTGGCGAGATTGCGCGAGCCTCTATCAAGATTGC 1437
442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1438 CCTCGGAGAGGCCACAGCTCAGGAAAGATGGTCATCATCACCAGGC 1487
459 roProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1488 CACCGGAAGCCAGTTCAAGGCCCGGAGGAGGATCTTTGGGAAATCGAA 1537
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisI 492
1538 GAGGAAACTTCTTTAACCCCAAGAGAAAGTGAAGCTGAAGGAGCATAT 1587

492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysT 509
|||||
1588 CAGAGTGCCTCTTCCACAGCTGGCCGGGTGATTGGCAAGGTGGCAAGA 1637
|||||
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValValProArg 525
|||||
1638 CCGTGAACGAACCTGCAGAACTTAACCACTGCAGAGTCATCGTGCCTCGT 1687
|||||
526 AspGlnThrProAspGluAsnAspGlnValValValValLysIleThrGlyH1 542
|||||
1688 GACCAACGCCAGATGAAATGAGGAAGTGTGTCAGATTATCGGGCA 1737
|||||
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
|||||
1738 CTTCTTTGCTAGCCACAGCTGCACAGCCGCAAGATCAGGGAAATTGTACAAC 1787
|||||
559 lnValLysGlnHisGlnGln 566
|||||
1788 AGGTGAACGACGAGGAGAGAAA 1810
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 22:11:59 ; Search time 93.87 Seconds
(without alignments)
902.223 Million cell updates/sec

Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKHQOQKALQSGPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_17:*
- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phase:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2956	100.0	579	4	O00425	O00425 homo sapien
2	2858	96.7	579	11	O9CPN8	O9cpn8 mus musculu
3	2483	84.0	593	13	O57526	O57526 xenopus lae
4	2461.5	83.3	592	13	O73932	O73932 xenopus lae
5	2261.5	76.5	584	13	O9PW80	O9pw80 brachydanio
6	2219.5	75.1	576	13	O42254	O42254 gallus gall
7	2192	74.2	577	11	O88477	O88477 mus musculu
8	2181	73.8	577	4	O9N218	O9n218 homo sapien
9	1868	63.2	556	4	O9V6M1	O9v6m1 homo sapien
10	835.5	28.3	558	5	O3VZ69	O3vz69 drosophila
11	835	28.2	566	5	O9NGS9	O9ngs9 drosophila
12	473	16.0	100	11	O9D054	O9d054 mus musculu
13	418.5	14.2	854	5	O21605	O21605 caenorhabdi
14	262.5	8.9	774	10	O9L128	O9l128 oryza sativ
15	243.5	8.2	768	5	O23487	O23487 caenorhabdi
16	240	8.1	644	4	O12828	O12828 homo sapien
17	238	8.1	641	5	O9BLA0	O9bla0 caenorhabdi
18	236.5	8.0	568	10	O9LXF5	O9lxf5 arabidopsis
19	235.5	8.0	621	10	O9C553	O9c553 arabidopsis

20	234	7.9	600	4	Q92946	Q92946 homo sapien
21	233	7.9	492	4	Q9UNW9	Q9unw9 homo sapien
22	233	7.9	498	4	O43267	O43267 homo sapien
23	231.5	7.8	313	10	O9L282	O9l282 arabidopsis
24	230.5	7.8	493	11	O9JKN6	O9jkn6 mus musculu
25	230	7.8	618	5	O9BLA1	O9bla1 caenorhabdi
26	229.5	7.8	398	3	O74919	O74919 schizosacch
27	229	7.7	386	5	P91632	P91632 drosophila
28	226.5	7.7	464	4	O9UEA1	O9uea1 homo sapien
29	226	7.6	1301	5	O9U982	O9u982 drosophila
30	226	7.6	1301	5	O9V8H6	O9v8h6 drosophila
31	225.5	7.6	577	10	Q9SRL3	Q9srl3 arabidopsis
32	225.5	7.6	589	5	Q17935	Q17935 caenorhabdi
33	225.5	7.6	611	5	Q17936	Q17936 caenorhabdi
34	222	7.5	479	10	O9XI71	O9xi71 arabidopsis
35	219.5	7.4	396	13	O9PSI1	O9psi1 xenopus lae
36	217.5	7.4	640	10	O9ASX3	O9asx3 arabidopsis
37	217.5	7.4	644	10	O9FNK3	O9fnk3 arabidopsis
38	215.5	7.3	403	11	O9DB01	O9db01 mus musculu
39	214	7.2	833	10	O9FMF0	O9fmf0 arabidopsis
40	213.5	7.2	680	5	P91393	P91393 caenorhabdi
41	213.5	7.2	711	4	O00301	O00301 homo sapien
42	213	7.2	721	11	O99PF5	O99pf5 rattus norv
43	211.5	7.2	561	5	O9GQN3	O9gqn3 drosophila
44	204.5	6.9	846	10	O49507	O49507 arabidopsis
45	202.5	6.9	1268	11	O9Z1A6	O9z1a6 rattus norv

ALIGNMENTS

RESULT 1
O00425 PRELIMINARY; PRT; 579 AA.
ID O00425
AC O00425;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE RNA BINDING PROTEIN KOC (KOC).
GN KOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL Oncogene 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U97188; AAC35208.1; -;
DR EMBL; U76705; AAD09223.1; -;
DR InterPro; IPR000504; RRM.
DR InterPro; IPR000958; KH.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS0030; RRM_RNP_1; FALSE_NEG.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
SQ SEQUENCE 579 AA; 63720 MW; AESC3A8EE3C135C5 CRC64;

Query Match 100.0%; Score 2956; DB 4; Length 579;
Best Local Similarity 100.0%; Pred No. 1.le-192;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPYSGPPLVKVTGYAFVDCPDSEWALKAIEALSGK 60
|||||
DB 1 MNKLYIGNLSENAAPSDLESIFKDAKIPYSGPPLVKVTGYAFVDCPDSEWALKAIEALSGK 60

```
QY 61 IELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
RL IELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
RN IELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
RP SEQUENCE FROM N.A.
RA Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RT Yuasa Y., Takeda M., Okano H.;
RL "Expression of mouse Igf2 mRNA-binding protein 3 and its implications
for the developing central nervous system.";
RT J. Neurosci. Res. 0:0-0(2001).
DR ENBL; AK011689; BAB27779.1; -.
DR ENBL; AB046173; BAB19755.1; -.
DR MGD; MGI:1919721; 261010IN1IRik.
DR InterPro; IPR000958; KH.
DR InterPro; IPR000504; RH.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 579 AA; 63574 MW; CABD9A4355B392B7 CRC64;

Query Match 96.7%; Score 2858; DB 11; Length 579;
Best Local Similarity 96.4%; Pred. No. 5.1e-186;
Matches 558; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNKLYIGNLSENAAPSDESIFKDAKIPVSGPFLVKTGYAFVDCPDDESALKATEALSGK 60
DB 1 MNKLYIGNLSDHAGPADLESVFDKAKIPVAGPFLVKTGYAFVDCPDDEGALKATEALSGK 60
QY 61 IELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
DB 61 MELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
QY 121 VVNTYSSKQARQALDKLNGFQLENFTLVKAYIPDDEMAAQNPLOQPRRGLGQRGSS 180
DB 121 VVNTYSSKQARQALDKLNGFQLENFTLVKAYIPDDEMAAQNPLOQPRRGLGQRGSS 180
QY 181 ROGSPGVSQKPCDPLRLLLVPTQFVGAIGKEGATIRNITKQTSKIDVHRKENAGAA 240
DB 181 ROGSPGVSQKPCDPLRLLLVPTQFVGAIGKEGATIRNITKQTSKIDVHRKENAGAA 240
QY 241 EKSITILSTPEGTSAAKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLGKEGRNLK 300
DB 241 EKSITILSTPEGTSAAKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLGKEGRNLK 300
QY 301 KIEQDQTKITISPLQELTLNPERTITVKGNETCAKAEEMKKIRSYENDIASMNL 360
DB 301 KIEQDQTKITISPLQELTLNPERTITVKGNETCAKAEEMKKIRSYENDIASMNL 360
QY 361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYQFEQSETVHVHOFIPALSVCAL 420
DB 361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYQFEQSETVHVHOFIPALSVCAL 420
QY 421 IGKQGHKIKOLSFPAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRYIGKIKEENFV 480
DB 421 IGKQGHKIKOLSFPAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRYIGKIKEENFV 480
QY 481 SPKEEVKLEAHIRVPSFAAGRVTKGKGTYNELONLSSAEVVPVPRDQTPDENQVVKIT 540
DB 481 SPKEEVKLEAHIRVPSFAAGRVTKGKGTYNELONLSSAEVVPVPRDQTPDENQVVKIT 540
QY 541 GHFYACQVQARKIQEILTQVKHQOQKALQSGPPQSRK 579
DB 541 GHFYACQVQARKIQEILTQVKHQOQKALQSGPPQSRK 579
```

```
RESULT 2
Q9CPN8 PRELIMINARY; PRT; 579 AA.
AC Q9CPN8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE 10 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:2610036B18, FULL INSERT SEQUENCE (IGF2 MRNA-BINDING PROTEIN 3).
GN 2610101N1IRIK OR MIMP3.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
```

```
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RT Yuasa Y., Takeda M., Okano H.;
RL "Expression of mouse Igf2 mRNA-binding protein 3 and its implications
for the developing central nervous system.";
RT J. Neurosci. Res. 0:0-0(2001).
DR ENBL; AK011689; BAB27779.1; -.
DR ENBL; AB046173; BAB19755.1; -.
DR MGD; MGI:1919721; 261010IN1IRik.
DR InterPro; IPR000958; KH.
DR InterPro; IPR000504; RH.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 579 AA; 63574 MW; CABD9A4355B392B7 CRC64;
```

Query Match 96.7%; Score 2858; DB 11; Length 579;
Best Local Similarity 96.4%; Pred. No. 5.1e-186;
Matches 558; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

```
QY 1 MNKLYIGNLSENAAPSDESIFKDAKIPVSGPFLVKTGYAFVDCPDDESALKATEALSGK 60
DB 1 MNKLYIGNLSDHAGPADLESVFDKAKIPVAGPFLVKTGYAFVDCPDDEGALKATEALSGK 60
QY 61 IELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
DB 61 MELHGKPIEVHSHVPRKQRIKRLQIRNIPPHLQWEVLDSLLVQYGVVESCQVNTDSETA 120
QY 121 VVNTYSSKQARQALDKLNGFQLENFTLVKAYIPDDEMAAQNPLOQPRRGLGQRGSS 180
DB 121 VVNTYSSKQARQALDKLNGFQLENFTLVKAYIPDDEMAAQNPLOQPRRGLGQRGSS 180
QY 181 ROGSPGVSQKPCDPLRLLLVPTQFVGAIGKEGATIRNITKQTSKIDVHRKENAGAA 240
DB 181 ROGSPGVSQKPCDPLRLLLVPTQFVGAIGKEGATIRNITKQTSKIDVHRKENAGAA 240
QY 241 EKSITILSTPEGTSAAKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLGKEGRNLK 300
DB 241 EKSITILSTPEGTSAAKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLGKEGRNLK 300
QY 301 KIEQDQTKITISPLQELTLNPERTITVKGNETCAKAEEMKKIRSYENDIASMNL 360
DB 301 KIEQDQTKITISPLQELTLNPERTITVKGNETCAKAEEMKKIRSYENDIASMNL 360
QY 361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYQFEQSETVHVHOFIPALSVCAL 420
DB 361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYQFEQSETVHVHOFIPALSVCAL 420
QY 421 IGKQGHKIKOLSFPAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRYIGKIKEENFV 480
DB 421 IGKQGHKIKOLSFPAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRYIGKIKEENFV 480
QY 481 SPKEEVKLEAHIRVPSFAAGRVTKGKGTYNELONLSSAEVVPVPRDQTPDENQVVKIT 540
DB 481 SPKEEVKLEAHIRVPSFAAGRVTKGKGTYNELONLSSAEVVPVPRDQTPDENQVVKIT 540
QY 541 GHFYACQVQARKIQEILTQVKHQOQKALQSGPPQSRK 579
DB 541 GHFYACQVQARKIQEILTQVKHQOQKALQSGPPQSRK 579
```

RESULT 3
O57526 PRELIMINARY; PRT; 593 AA.
ID O57526
AC O57526
DT 01-JUN-1998 (TEMBLrel. 06, Created)


```
QY 351 YENDIASNNQAHLIPGLNINLGLFP-PTSGMPPPTSGPPS-AMTPPYQF-EQSTET 407
Db 361 YENDIAAMNQAHLIPGLNINLGLFPSSSGMPPPPSVGVPSTSTSYPPFGQOPESET 420
QY 408 VHOFPALSGAIIKGOGHIKOLSRFAGASIKIAPAEAPDAKVRWVIITGPPPEAQK 467
Db 421 VHLFPALAVGAIIKGOGHIKOLSRFAGASIKIAPAEAPDAKVRWVIITGPPPEAQK 480
QY 468 GRIYGIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVPDQ 527
Db 481 GRIYGLKEENFPGPREVKLEHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVPDQ 540
QY 528 TPDENDVVYKITGHFYACQVQARKIQEILITQVK--QHQQOKALQSGPPPSRRK 579
Db 541 TPDENDVVYKITGHFYACQVQARKIQEILITQVK--QHQQOKALQSGPPPSRRK 594

RESULT 5
Q9PW80 PRELIMINARY; PRT; 582 AA.
AC Q9PW80;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Vcl RNA BINDING PROTEIN.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasboridae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA Taylor W., Meyer D., Standart N., Raz E., Visraeli J.K.;
RT "vcl RBP intracellular distribution and evolutionarily conserved
RT expression suggest multiple roles during development.";
RL Mech. Dev. 0:0-0(1999).
DR EMBL: AF161270; AAD45610.1; -.
DR InterPro: IPR000504; RRM.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS50102; RRM; 2.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
SQ SEQUENCE 582 AA; 63351 MW; 9DA563200681B306 CRC64;

Query Match 76.5%; Score 2261.5; DB 13; Length 582;
Best Local Similarity 77.1%; Pred. No. 1.6e-145;
Matches 450; Conservative 50; Mismatches 77; Indels 7; Gaps 5;

QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSGK 60
Db 1 MNKLYIGNVSEASALDESIFEQWIKIPSPAPFLVSGYAFVDCPDSEWALKAEALDLSGK 60
QY 61 IELHGKPIEVHSPVKRQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
Db 61 VELHGVLEHVSVPKRRQRIRKLIQIRNIPPHMQWEVLDLLGALQYGVSECEQVNTDSETA 120
QY 121 VNVVYSSKDAQRAQDLKNGFQLENFTLKVAYIPDEMAAQNPLOQPRRRGLGQRSS 180
Db 121 VNVVRYGAKDAQRAEAMDKNGFLMENVALKVSIPTETAADAP--AVGRRGPNRGP 178
QY 181 ROGSPGSVKQK--PCDLRLVPTQFVGAIIGKEGATIRNITKQSKIDVHRKENA 239
Db 179 RQSPSLGAKRPLQSDVPLRLVPTQFVGAIIGKEGATIRNITKQTHSIDHRKENA 238
QY 240 AKSITILSTPEGTSACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRLLIGKEGRNL 299
Db 239 AKRPITVHSTPEGSSACRNIMEQKEADTKITEIPLKILAHNNFVGRLLIGKEGRNL 298
```

```
QY 300 KKIEQDTRDTKITITISPLQELTYLNPRTITVKNVETCAKAEIEIMKKIRSEYENDIASMN 359
Db 299 KKIEQDTRDTKITITISPLQELTYLNPRTITVKTLDCAKAEIEIMKKVRESYENDVAAMH 358
QY 360 LOAHLIPGLNINLGLFP-PTSG--MPPTSGPPSAMTPPYQF-EQSETETVHQFIPAL 415
Db 359 LQSNLIPGLNINLGLFPGAASGGSISVSVGGPPGGAAGYQSGAQMSETHLFLIPAL 418
QY 416 SVGAIIKGOGHIKOLSRFAGASIKIAPAEAPDAKVRWVIITGPPPEAQKAGRIYGIK 475
Db 419 AVGAIIKGOGHIKOLSRFAGASIKIAPADIDAKVRWVIISGPPPEAQKAGRIYGIK 478
QY 476 EENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVPDQTPDENDQV 535
Db 479 EENFFGPKKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVPDQTPDENDQV 538
QY 536 VVKITGHFYACQVQARKIQEILITQVKHQOQOKALQSGPPPSRRK 579
Db 539 VVKITGHFYACQVQARKIQEILITQVKHQOQOKPKPSAAGPPVARRK 582

RESULT 6
Q42254 PRELIMINARY; PRT; 576 AA.
AC Q42254;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ZIPCODE-BINDING PROTEIN.
GN ZBPL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97220007; PubMed-9121465;
RA Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL Mol. Cell. Biol. 17:2158-2165(1997).
DR EMBL: AF026527; AAB82295.1; -.
DR InterPro: IPR000504; RRM.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS50102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
SQ SEQUENCE 576 AA; 63271 MW; 01AAF2D1D81C8811 CRC64;
```

Query Match 75.1%; Score 2219.5; DB 13; Length 576;
Best Local Similarity 74.7%; Pred. No. 1.2e-142;
Matches 437; Conservative 61; Mismatches 72; Indels 15; Gaps 7;

```
QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKAEALSGK 60
Db 1 MNKLYIGNLSNESVTPADLEKVFNDHKISFSGQLVKSQYAFVDCPDSEWALKAEALSGK 60
QY 61 IELHGKPIEVHSPVKRQRIRKLIQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
Db 61 VELHGVLEHVSVPKRRQRIRKLIQIRNIPPHLQWEVLDLLGALQYGVVSECEQVNTDSETA 120
QY 121 VNVVYSSKDAQRAQDLKNGFQLENFTLKVAYIPDEMAAQNPLOQPRRRGLGQRSS 179
Db 121 VNVVYTNREQTQATMKLNHQLNHLVKVSIPTESQSVQ----GPNERRGGFGARGA 176
QY 180 ROGSP---GSVSKQPCDLRLVPTQFVGAIIGKEGATIRNITKQSKIDVHRKEN 236
Db 177 PROGSPVYAPVQKQPDVPLRLVPTQVVGAIIGKEGATIRNITKQSKIDVHRKEN 236
```


Qy 237 AGAAEKSTILSTPECTSAACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRIGKEG 296
 Db 237 AGAAEKAIHSITPEGCSAAKMLEIMHKEAKDTKTADVEPLKILAHNNFVGRIGKEG 296
 Qy 297 RNLLKIEODTQTKITISPLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Db 297 RNLLKIEODTQTKITISPLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Qy 357 SMNLQAHILPGLNLAALGFPPTSGMPPPTSGPPSAMT--PPYQFQESQETVHQFIPA 414
 Db 357 AMSLQSHLIPGLNLAALGFPASSNAVPP---PPSSVGAAPYSFMPPEQETVHVFI 413
 Qy 415 LSVGAIGKQGHKQILSRFAGASIKIAPAEAPDAKVRWVITGPPEAQFKAQRIYOKI 474
 Db 414 QAVGAIGKQGHKQILSRFASIKIAPPETPDSKVRWVITGPPEAQFKAQRIYOKL 473
 Qy 475 KEENFVSPKEVKLEAHLRVPFSAAGRVIGKGTVNELQNLSSAEVVVPRDQTPDENQ 534
 Db 474 KEENFVSPKEVKLEAHLRVPFSAAGRVIGKGTVNELQNLSSAEVVVPRDQTPDENQ 533
 Qy 535 VVVKITGHFYACQVAQRKIQIBLTQVKHQHQQKALQSGPPQSRK 579
 Db 534 VIVKIGHFYASQMAQRKIRILAQVKQ-QHKG-QSGQLQARRK 576

RESULT 7
 O88477 PRELIMINARY; PRT; 577 AA.
 AC O88477;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CODING REGION DETERMINANT BINDING PROTEIN (CODING REGION DETERMINANT-
 BINDING PROTEIN).
 GN CRDBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92217743; PubMed=1559612;
 RA Bernick P.L., Herrick D.J., Prokipcak R.D., Ross J.;
 RT "Control of c-myc mRNA half-life in vitro by a protein capable of
 RL binding to a coding region stability determinant.";
 RL Genes Dev. 6:642-654(1992).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=9415886; PubMed=8114742;
 RA Herrick D.J., Ross J.;
 RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
 RL influence of the coding and 3' untranslated regions and role of
 RT ribosome translocation.";
 RL Mol. Cell. Biol. 14:2119-2128(1994).
 [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=94179348; PubMed=8132663;
 RA Prokipcak R.D., Herrick D.J., Ross J.;
 RT "Purification and properties of a protein that binds to the C-terminal
 RL coding region of human c-myc mRNA.";
 RL J. Biol. Chem. 269:9261-9269(1994).
 [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=97322234; PubMed=9178888;
 RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
 RA Gruppiso P.A., Ross J.;
 RT "Developmental regulation of CRD-BP, an RNA-binding protein that
 RL stabilizes c-myc mRNA in vitro.";
 RL Oncogene 14:1279-1286(1997).
 [5]
 RN SEQUENCE FROM N.A.
 RP Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
 RA Fleisig A.J.

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AF061569; AAC72743.1;
 DR EMBL; AK013940; BAB29071.1;
 DR MGD; MGI:1330862; Cribp.
 DR InterPro; IPR000958; KH.
 DR InterPro; IPR000504; RH.
 DR Pfam; PF00013; KH-domain; 4.
 DR Pfam; PF00076; rim; 2.
 DR SMART; SM00322; KH; 4.
 DR SMART; SM00360; RM; 2.
 DR PROSITE; PS0102; RM; 2.
 SO SEQUENCE 577 AA; 63450 MW; EFBIAF2FF9F0344 CRC64;

Query Match 74.2%; Score 2192; DB 11; Length 577;
 Best Local Similarity 74.1%; Pred No. 8.5e-141;
 Matches 434; Conservative 62; Mismatches 74; Indels 16; Gaps 8;

Qy 1 MNKLYIGNLSENAAPSDESIFPKDAKIPVSGPLFKVTGYAFVDCPDSEWALKAEIALSK 60
 Db 1 MNKLYIGNLSENAAPSDESIFPKDAKIPVSGPLFKVTGYAFVDCPDSEWALKAEIALSK 60
 Qy 61 IELHCKPIEVHSVPKRIKQIRNIPPHLQWELVDSLLVQYGVSCQVNDSETA 120
 Db 61 VELOKRLIEIHSVPKRSRKIRNIPPHLQWELVDSLLVQYGVSCQVNDSETA 120
 Qy 121 VVNVTVSKDQARQALDKLNGFLENFTLVKVAIPDEMAAQNPQQPRRR-GLGQSGS 179
 Db 121 VVNVTVSKDQARQALDKLNGFLENFTLVKVAIPDEMAAQNPQQPRRR-GLGQSGS 179
 Qy 180 SRQSGSP---GSVSKQKPCDPLRLVLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKN 236
 Db 180 SRQSGSP---GSVSKQKPCDPLRLVLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKN 236
 Qy 237 AGAAEKSTILSTPECTSAACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRIGKEG 296
 Db 237 AGAAEKSTILSTPECTSAACKSILEIMHKEAQDIKFTTEIPLKILAHNNFVGRIGKEG 296
 Qy 297 RNLLKIEODTQTKITISPLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Db 297 RNLLKIEODTQTKITISPLQELTYNPERTTVKGNVETCAKAEEMKIKRESYENDIA 356
 Qy 357 SMNLQAHILPGLNLAALGFPPTSGMPPPTSGPPSAMT--PPYQFQESQETVHQFIPA 413
 Db 357 AMSLQSHLIPGLNLAALGFPASSNAVPP---PPSSVGAAPYSFMPPEQETVHVFI 413
 Qy 414 LSVGAIGKQGHKQILSRFAGASIKIAPAEAPDAKVRWVITGPPEAQFKAQRIYOKI 473
 Db 414 QAVGAIGKQGHKQILSRFASIKIAPPETPDSKVRWVITGPPEAQFKAQRIYOKL 473

```
QY 474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 LKEENFPGKPEEVKLETHIRVPASAAAGRVIGKGGKTVNELQNLTAEEVVVPRDQTPDEND 533

QY 534 QVVVKITGHFYACOVAQRKIQTIELTVQVKHQHQQKALQSGPPQSRKK 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 534 QVIVKIIGHFYASOMAQRKIRDLAQVKQ-OHQKG-QSNLAQARRK 577

RESULT 8
Q9NZI8 PRELIMINARY; PRT; 577 AA.
ID Q9NZI8
AC Q9NZI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MRNA-BINDING PROTEIN CRDBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA Panoutsakopoulos G., Kyriazoglou I., Voutzoulis S., Tsiapalis C.M.,
RA Kittas C., Agnantis N., Pandis N.;
RT "Ectopic expression of a KH-domain containing protein, highly
RT homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT malignant mesenchymal tumors.";
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF198254; AAF37203.1; -.
DR InterPro: IPR000958; KH.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS50102; RRM; 2.
SQ SEQUENCE 577 AA; 63456 MW; 0749A060F252D81D CRC64;

Query Match 73.8%; Score 2181; DB 4; Length 577;
Best Local Similarity 73.9%; Pred. No. 4.8e-140;
Matches 433; Conservative 61; Mismatches 76; Indels 16; Gaps 8;

QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPLVKTGYAFVDCPDSEWALKAIETLSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MNKLYIGNLSENVTPADLEKVFPAEHKISYSGQFLVKSQYAFVDCPDSEWAMKAETFSGK 60

QY 61 IELHGKPIEVESHVSPKQRIRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VELQKRLEIHSVPKQSRKQIRNIPPLRWELVDSLLAQYGVNCEQVNTDSETA 120

QY 121 VNVVTSYSSKDOARQALDKLNGFLENFTLKVAYIPDEMMAQNPLOQPRGRGLGQSGS 179
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 VNVVTSYNSREQTQATMKLNGHLENHALKVSYPDEQIAQ-----GPENGRGGFGSGRG 176

QY 180 SROGSP---GSVSKQKPCDPLRLVPTQFVGAIIGKEGATIRNITQTSKIDVHRKEN 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 PROGSVAAGAPAKQOQVQDIPRLVPTQYVGAIIGKEGATIRNITQTSKIDVHRKEN 236

QY 237 AGAAEKSIITLSTPEGTSACKSILEIMHKEADIKFTEIPLKILAHNNFVGRLLIGKEG 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 237 AGAAEKASIVHSHTPEGSSACKWILEIMHKEADTKTAEVPLKTLAHNNFVGRLLIGKEG 296

QY 297 RNLKKEIQDFTKTIITSPLOELTYLNPRTITVKGNVETCAKAEIEIMKKIRESYENDIA 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 RNLKKEIQDFTKTIITSSLODLYLNPRTITVKGAIENCCRAEQEIMKKVREAYENDVA 356

QY 357 SNLQAHLLPGLNLNALGLFPPTSGMPPTSGPPSAMT--PPYPQEQS-ETETVHQFTIP 413
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 357 ANSLQSHLLPGLNLAAVGLFPASSSAVPP---PPSSVTGAAPYSPPMQAPEQEMVQVFTIP 413
```

```
QY 414 ALSVGAIIGKQGHQIKOLSRFAGASIKIAPAEAPDAKVRMVIITGPPPEAQKAGRIYGK 473
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 AQAVGAIIGKGGHIGKLSRFASASIKIAPETPDSKVRMVIITGPPPEAQKAGRIYGK 473

QY 474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 LKEENFPGKPEEVKLETHIRVPASAAAGRVIGKGGKTVNELQNLTAEEVVVPRDQTPDEND 533

QY 534 QVVVKITGHFYACOVAQRKIQTIELTVQVKHQHQQKALQSGPPQSRKK 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 534 QVIVKIIGHFYASOMAQRKIRDLAQVKQ-OHQKG-QSNLAQARRK 577

RESULT 9
Q9Y6M1 PRELIMINARY; PRT; 556 AA.
ID Q9Y6M1
AC Q9Y6M1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HEPATOCELLULAR CARCINOMA AUTOANTIGEN.
GN P62.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99207072; PubMed-10190901;
RA Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT in human hepatocellular carcinoma.";
RL J. Exp. Med. 189:1101-1110(1999).
DR EMBL; AF057352; AAD31596.1; -.
DR InterPro: IPR000504; RRM.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS50102; RRM; 2.
DR SMART: SM00322; KH; 4.
DR SMART: SM00360; RRM; 2.
SQ SEQUENCE 556 AA; 61842 MW; 1EDEFB100443DDC4 CRC64;

Query Match 63.2%; Score 1868; DB 4; Length 556;
Best Local Similarity 64.8%; Pred. No. 8.1e-119;
Matches 368; Conservative 78; Mismatches 96; Indels 26; Gaps 6;

QY 1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPLVKTGYAFVDCPDSEWALKAIETLSGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 MNKLYIGNLSPAVTADDLRLQFGDKRLPLAGQVLLKSGYAFVDPDQNWAIETLSGK 61

QY 61 IELHGKPIEVESHVSPKQRIRKQIRNIPPHLQWEVLDLLVQYGVVSECEQVNTDSETA 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 VELHKGIMEVDYSVSKLSRKQIRNIPPHLQWEVLDGLLAQYGVNVEQVNTDSETA 121

QY 121 VNVVTSYSSKDOARQALDKLNGFLENFTLKVAYIPDEMMAQNPLOQPRGRGLGQSGS 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 VNVVTSYATREEAKIAMEKLSGHQFENYFKIYIPDEEVSSPSPQ--RAQR--GDHSR 177

QY 181 ROG-SPGVSQKQPCDPLRLVPTQFVGAIIGKEGATIRNITQTSKIDVHRKENAGA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 EQHAFEGGTQSOARQIDFPLRLVPTQFVGAIIGKEGLTKNITQTSRVDIHRKNSGA 237

QY 240 AKSITILSTPEGTSACKSILEIMHKEADIKFTEIPLKILAHNNFVGRLLIGKEGRNL 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 AEPKPTIHATPEGTSEACRMILEIMQKEADETKLAEIPLKILAHNGLVGRLLIGKEGRNL 297

QY 300 KKEQDFTKTIITSPLOELTYLNPRTITVKGNVETCAKAEIEIMKKIRESYENDIASMN 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 KKEHETGKTIITSSLODLYLNPRTITVKGTVCAKAEIEIMKKILREAFENDMLAVN 357
```


[illegible]

```

RESULT 14
Q9LI28
ID Q9LI28 PRELIMINARY; PRT; 774 AA.
AC Q9LI28;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ESTS D23839(R0339).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Magnoliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RC Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:p0708G02.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP001539; BA92910.1; -.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 4.
DR

```

RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Wieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinon-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: U00047: AAA50693.1; -;
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR000536; Hormone_rec_lig.
DR InterPro: IPR000958; KH.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00013; KH-domain; 4.
DR Pfam: PF00104; hormone_rec; 1.
DR PROSITE: PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00322; KH; 4.
SQ SEQUENCE 768 AA; 83886 MW; 3672FC807710F20A CRC64;

Query Match 8.2%; Score 243.5; DB 5; Length 768;
Best Local Similarity 21.2%; Pred. No. 2.6e-08;
Matches 123; Conservative 103; Mismatches 215; Indels 139; Gaps 22;
QY 62 ELHGKPIEVEHSPKQRIKRLQIRNIPPHLQWEVLDSLQYGVVESCQVNTDSETAV 121
DB 139 QLICKLIELD--ITKNESL-----LSALFL-----CSAVHPDISTEG 174
QY 122 VNTYSSKQDQAQAL-----DKLNGFQ--LENFTLKVAYIPDEMAAQON 163
DB 175 KNLLYKYYQVYSAALINHCCLTNQONAPRSYSELLSYQIIEETHQKIIIGADGGIGTLKR 234
QY 164 PLQO-----PRGRGLQGRSSRQSPGSVSKQKPCDPLPLLLVPTQFVGAIGKEG 215
DB 235 PLQSEILDGLIPTKKS--SEVGLNMGDSKDTDIYP-----VPEKVVGLVIGKG 284
QY 216 ATIRNITKQTSKIDVHRKENAGAAEKSIITLSTPETSAAKSIILEIMHKE---AODIK 272
DB 285 SEIRLIQOTSQCRVOMDPDQHSVNGFNCTIEGPPDQAVAROMITQVINRQTGAQPGA 344
QY 273 FTEEIPKILAHNNFVGRLLGEGRNKKTEQDITDKITISPLQELTLY----NPERTIT 328
DB 345 APGEVTEMLIPADKIGLVIGKGETIRIVQEQSGLR--NCNVVQETTATGQPKPLRMIG 403
QY 329 VGNVETCAKAEIEIMKKIRESYENDIASMNLQAHLLIPGLNLAGLFPPTSGMPPTSG 388
DB 404 SPAAIET-AKA-----LVHNMNNTQGNAPLLQRAHPQPSG 438
QY 389 PPSAMTPPYQF-----EQSETETVHQFIPALSVGAILGKOGOHIKOLSRFAGASIK 440
DB 439 -----QFGGGYCAQEAQAKGEVI---VPLRSLAGMIIGKGGEMIKRLAAETGTGIQ 485
QY 441 IAPAEAPDAKRVVITGPPEAFKAGRIYGIKE--ENFVSPKEVLEA----- 490
DB 486 FKPDTPNSEDRIAVINGTRDQIYRATERITEIVNRAIKNNGAPQDRGSAGTVLPQOSIF 545
QY 491 HIRVPSFAAGRGVIGKGTVNELQNLSAEV--VVRDQTPDENDQVVVKITGH-----FYA 545
DB 546 YMHVPAGKCGLVIGKGENIKITERTGTATCGLAPAAEQKNE--DEKVFEEKSGLQIHA 604
QY 546 CQVAQRKIQEILTQVK-----QHQQOKALQSGPPQS 576
DB 605 SHLVRIKVGELSPNTVPVPLQAGAGGYQQQQAAMFSAQTQN 644

Search completed: January 25, 2002, 22:19:19
Job time: 440 sec

OM of: US-09-685-696-176 to: Issued_Patents_NA:* out_format : pfs
Date: Jan 25, 2002 11:22 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-O=/cgn2_1/uspt0.spool/us09685696/runat_25012002.145551.10660/app_query.fasta_1.648
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -XGAPEXT=0.500
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=2000000000
-USER=us09685696 -CGN1_1_87 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-685-696-176
Query length: 579
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 11323899
Search time (sec): 90.080000

score_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-4	2956.00	5412.74	7.7e-294	4159	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-261-855-1	2192.00	4009.61	1.1e-215	2224	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-6	1950.50	3538.22	1.5e-190	3412	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-7	1873.00	3422.14	5.8e-183	1946	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-8	1868.00	3406.34	4.4e-182	1728	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-5	1637.00	2987.92	8.9e-159	1708	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-1	241.00	405.53	6.1e-15	2384	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-1	241.00	405.53	6.1e-15	2384	1
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01782-1	241.00	405.53	6.1e-15	2384	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-9	236.50	397.23	1.8e-14	2381	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-9	236.50	397.23	1.8e-14	2381	1
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01782-9	236.50	397.23	1.8e-14	2381	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-7	229.50	387.80	6.0e-14	1803	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-7	229.50	387.80	6.0e-14	1803	1
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01782-7	229.50	387.80	6.0e-14	1803	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-187-793-3	227.00	375.52	2.9e-13	3318	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-866-928B-2	159.50	261.55	6.4e-07	1416	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-613-182-5	159.50	252.78	2.0e-06	2846	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-187-793-1	154.00	243.03	6.9e-06	2756	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-021-608D-5	147.00	241.67	8.2e-06	1097	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-5	147.00	241.67	8.2e-06	1097	1
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-01782-5	147.00	241.67	8.2e-06	1097	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-667-276A-3	145.50	231.25	3.1e-05	2017	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-347-833-3	140.50	220.27	0.0001	2318	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-973-273-2	134.50	199.26	0.0019	5113	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-881-075-50	133.00	212.16	0.0004	1467	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-120-827-50	133.00	212.16	0.0004	1467	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-478-675-50	133.00	212.16	0.0004	1467	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-461-027-1	133.00	212.15	0.0004	1469	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-082-501-1	133.00	212.15	0.0004	1469	1
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US92-01836-1	133.00	212.15	0.0004	1469	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-990-114-2	132.50	206.86	0.0007	2079	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-935-450-5	132.00	202.73	0.0012	2685	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-343-011B-4	128.50	200.63	0.0016	1896	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-702-771-2	128.00	202.78	0.0012	1485	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-438-883-2	128.00	202.78	0.0012	1485	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-741-599-2	128.00	202.78	0.0012	1485	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-347-833-1	127.00	194.35	0.0036	2508	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-967-4	121.50	185.40	0.0112	2277	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-967-5	121.50	185.40	0.0112	2277	1

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-974-4 + 121.50 185.40 0.0112 2277
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-974-5 + 121.50 185.40 0.0112 2277
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-098-487-4 + 121.50 185.40 0.0112 2277
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-098-487-5 + 121.50 185.40 0.0112 2277
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-676-967-3 + 121.50 183.11 0.0150 2733
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-4
seq_documentation_block:
; Sequence 4, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 4
; LENGTH: 4159
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-4

alignment_scores:
Quality: 2956.00 Length: 579
Ratio: 5.105 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-685-696-176 x US-09-061-709-4
Align seg 1/1 to: US-09-061-709-4 from: 1 to: 4159
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAProSerAs 17
|||||
251 ATGAACAACTGTATATCGAAACCTCAGCAGAACGCCGCCCTCGGA 300
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
301 CCTAGAAAGTATCTCAAGAGCGCCCAAGATCCCGTCTCGGACCTTCC 350
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
351 TGGTGAACACTGGCTTACCGCTTCGGACTGCCCGCAGAGCTGGGCC 400
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
401 CTCAAGGCCATCAGGCGCTTTCAGSTAAATAGAACTGCACGGGAACC 450
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
451 CATAGAAGTTGAGCACTCGTCCCAAAAGGCAAGAGATTCCGAACATTC 500
84 InIleArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
501 AGATACGAAATATCCCGCTCATTTACAGTGGGAGGTCTGGATAGTTTA 550
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
551 CTAGTCCAGTATGAGTGGTGGAGACCTGTGAGCAAGTGAACACTGACTC 600
117 rGluThrAlaValValAsnValThrTyrSerLysAspGlnAlaArgG 134
|||||

```
601 GGAAACTGCAGTTGTAATGTAACCTATTTCAGTAAGGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
651 AAGCACTAGACAAACTGAATGATTTTCAGTTAGAGAAATTTACCTTGAAA 700
|||||
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG1 167
|||||
701 GTAGCCTATATCCCTGATGAATAGCCGCCAGCAAAACCCCTTCAGCA 750
|||||
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlys 184
|||||
751 GCCCGAGTCCCGGGGCTGTGGCAGAGGGGCTCCTCAGGACAGGGGT 800
|||||
184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
801 CTCCAGGATCCGTATCCCAAGCAGAAACCATGTGATTTGCCCTCTGCCCTG 850
|||||
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
|||||
851 CTGGTTCCACCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 900
|||||
217 rIleArgAsnIleThrLysGlnThrGlnPheLysIleAspValHisArgL 234
|||||
901 CATTCGGAACATCATCAACACAGACCCAGTCTAAATCGATGTCACCGTA 950
|||||
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
|||||
951 AAGAAATGCGGGGCTGCTGAGAAATCGATTACTATCTCTACTCCT 1000
|||||
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG1 267
|||||
1001 GAAGCACCTCTCGCGCTGTGAAGTCTATTCTGGAGATTATGCAATAAGGA 1050
|||||
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
1051 AGCTCAAGATATAAATTCACAGAGAGATCCCTTGGAATTTAGCTC 1100
|||||
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
1101 ATAATACTTTGTTGGAGCTCTTATTGTTAAAGAGGAAGAAATCTTAAA 1150
|||||
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG1 317
|||||
1151 AAAATTGAGCAAGACACAGACACATAAATCACGATATCTCATTCAGAGA 1200
|||||
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
1201 ATTGACGCTGTATATCCAGAACCGCACTATTACAGTTAAAGGCAATGTTG 1250
|||||
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1251 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGAGTCT 1300
|||||
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG1 367
|||||
1301 TATGAAAATGATATGCTTCTATGAATCTCAAGCACATTTAATTCCTGG 1350
|||||
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1351 ATTTAAATCTGAACGCTTGGGTCTGTTCCACCCACTTCAGGGATGCCAC 1400
|||||
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1401 CTCCACCTCAGGGGCCCTTTTCAGCATGACTCTCCCTACCCGCAAGTTT 1450
|||||
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1451 GAGCAATCAGAAACGGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
|||||
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
|||||
1501 CGGTGCCATCATCGCAAGCAGGCGCCAGCACATCAAGCAGCTTCTCGCT 1550
```

```
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1551 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAAGCACAGATGCTAAA 1600
|||||
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG1 467
|||||
1601 GTGAGGATGGTGATTATCAGTGACGACAGAGGCTCAGTTCAAGGCTCA 1650
|||||
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1651 CGGAAGAATTTATGCAAAATTAAGAGAGAAACTTTGTTAGTCTCTAAG 1700
|||||
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaLaGly 500
|||||
1701 AAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCATCTCTTCTGCTGCG 1750
|||||
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1751 AGAGTTATTGGAAGAGGAGGCAAAACGGTGAATGAACCTCAGAATTTGTC 1800
|||||
517 rSerAlaGluValValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1801 AAGTCACAGAAGTTGTTGCTCCCTGTCGACGACACACTGATGAGAATGACC 1850
|||||
534 InValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1851 AAGTGGTTGTCAAAATAACTGGTCACTTCTATGCTTGCCAGGTTGCCAG 1900
|||||
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLys 567
|||||
1901 AGAAAATTCAGGAAATTTCTGACTCAGGTAAGCAGCACCACACAGAA 1950
|||||
567 saAlaLeuGlnSerGlyProGlnSerArgArgLys 579
|||||
1951 GGCCTCGAAAGTGGACCACCTCAGTCAAGACGGAG 1987
|||||
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-261-855-1

```
seq_documentation_block:
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-261-855-1
```

```
alignment_scores:
Quality: 2192.00      Length: 585
Ratio: 4.152          Gaps: 8
Percent Similarity: 90.256      Percent Identity: 74.188

alignment_block:
US-09-685-696-176 x US-09-261-855-1

Align seg 1/1 to: US-09-261-855-1 from: 1 to: 2224

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
131 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGTGTCGCCCGCAG 180
|||||
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
```

```

181 CTTGGAGAAAGTATTCCGCGAGCACAAAGATCTCTACAGCGCCAGTTCT 230
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
231 TGGTCAATCCGGCTACCGCTTCGTGGATTGCCCGACGAGCACTGGCG 280
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
281 ATGAAGCCATCGAAATTTCTCGGGAAAGTAGAACTCGAAGAAACG 330
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
331 TCTAGAGATTGAACACTCAGTCCCAAAACAAAGAGAGTCCGAAATAC 380
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
381 AGATCCGCAATATTCCACCTCAGCTCCGATGGGAAGTGTAGATACCTG 430
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
431 CTGGCTCAGTACGGTACAGTGGAGAACTGTGAGCAAGTGAACACTGAA 480
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
481 TGAGACAGCGGTGGTCAAGCTCACCCTACTCTAACCGGGAGCAGACCG 530
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
531 AGCTATCATGAAGCTAAATGGCCATCAACTGGAGAACCATGCCCTGAAG 580
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
581 GTCTCTACATACCTGATGAGCAGATAACA...CAAGGTCTCAGATG 627
167 nProArgGlyArgArgGlyLeuGlnArgGlySerSerArgGlnGlyS 184
628 GCGTCGTGGA.....GGCTTTGGTCTCGGGGCCAGCCCGCAAGGT 671
184 exPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
672 CGCCCTGGCAGCAGGGGCTCCAGCCAAAGCAGCAGCCAGTGGACATCC 721
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGl 214
722 CTCCGGCTCTCGTGGCTACGCGAGTATGTAGCGCTATCATTTGGCAAG 771
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
772 GGTGGCCACCATCCGAACATCACAAACAGACGCGCAGTCCAAATAGAC 821
231 aHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
822 TGCAATAGGAAGAGAAATCGGGCGCTCGCGAAGGCCATCAGCGTCAT 871
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIle 264
872 TCAACCCCTGAAGCTGCTCTCCCGCTGCAAGATGATCTTGGAGATTAT 921
264 tHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
922 GCACAAGGAGGCAAGGACACCAACACCGCAGATGAAGTTCCCTCGAAG 971
281 leuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
972 TCTGGCTCATACAACTTCGTGGCGCACTCATTTGGCAAGGAGCGCG 1021
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
1022 AACCTGAAGAGGTGGAGCAGACACAGAGACAGATCACCATCTCATC 1071
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331

```

```

1072 GTCCAGGACCTCACGCTCTATAACCTTGAGAGACCATCACATGTGAAG 1121
331 lyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIle 347
1122 GCCCCAATTGAGAACTGTTGAGGGCCGAGGAGGATCATGAAGAAGTT 1171
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
1172 CGAGAGGCTTACGAGAACGACGTGGCGCCATGAGCTTGAGTCCACCT 1221
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerG 381
1222 CATCCCTGGGCTTAACCTGCTGCTAGTCTCTCCAGCTTCATCCA 1271
381 lyMetProProProThrSerGlyProProSerAlaMetThr.....Pro 395
1272 GCGCTGCTCCTCT.....CTCCCAAGCAGTGTCACTGGGGCTGCT 1312
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
1313 CCCTATAGCTCTTCATGCGAGCTCCGAGCAGGATGGTACAAGTGT 1362
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisI 428
1363 CATCCCGCGCCAGGCTGTGGCGCCATCAATTGGCAAGAGGCCACACA 1412
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
1413 TCNAACAACCTCTCCGTTCCGAGCGCTCCATCAAGATTGCTCCACA 1462
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProGl 461
1463 GAAACACCTGACTCCCAAGTTCGAATGGTCTCATCACTGACCCCCAGA 1512
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluA 478
1513 GGCTCAGTTCAAGGCTCAGGAAAGAAATTTATGGCAAACTAAAAAGAAGA 1562
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
1563 ATTTCCTTGGTCCCAAGGAGAAAGTAAAGTAGAGCCACATACGGGTT 1612
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValas 511
1613 CCGGCTTCAGCAGCCGCGCTCATCGCAAAAGCGGCAAAACGGTGAA 1662
511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
1663 TGAGCTGCAGAACTTGAAGTTCAGCTGAGGTGGTAGTGCCCAAGAGACCA 1712
528 hrProAspGluAsnAspGlnValValLysIleThrGlyHisPheTyr 544
1713 CCCCGATGAGAAACGACCAAGTCAATGTTAAGATCATCGGACATTTCTAT 1762
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVally 561
1763 GCCAGCAGATGGCTCAGCGGAAGATCCGAGACATCTCGCTCAAGTTAA 1812
561 sGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
1813 GCAA...CAGCACCAAGGGA...CAGAGCAACCTGGCCCGCAGGACGGA 1856
578 rglys 579
1857 GGAAG 1861

```

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-6
seq_documentation_block:
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng

```

; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-6

alignment_scores:
Quality: 1950.50 Length: 591
Ratio: 3.817 Gaps: 8
Percent Similarity: 86.464 Percent Identity: 65.313

alignment_block:
US-09-685-696-176 x US-09-061-709-6
Align seg 1/1 to: US-09-061-709-6 from: 1 to: 3412

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
73 ATGAACAAGCTTTACATCGGAACCTTGAGCCCGCCGCTCACCGCGGACGA 122

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
123 CCTCGGAGCTCTTTGGGACAGGAAGCTGCCCTGGCGGACAGGTCC 172

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
173 TGCTGAAGTCGGGTACGCCCTTCGTGGACTACCCCGACCAAGCTGGGCC 222

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
223 ATCCGCGCATCGAGACCTCTCGGGTAAGTGAATTCATGGGAAAT 272

67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
273 CATGGAAGTTGATTACTACTGCTCTTAAAGCTAAGGAGCAGGAATTC 322

84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
323 AGATTGCAACATCCCTCCTCACCTGCAGTGGAGGTGTTGGATGGACTT 372

101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
372 TTGGCTCAATATGGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAC 422

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
423 AGAACCCTGCTTCAACGTCACATATGCAACAAGAGAAGAACAAAAA 472

134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
473 TAGCCATGGAGAAGTAAGCGGCATCAGTTTGAGAACTACTCTTCAAG 522

151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnGI 167
523 ATTTCCCTACATCCCGATGAAGAGGTGAGTCCCTTCGCCCCCTCAG.. 570

167 nProArgGlyArgGlyLeuGluGlnArgGlySerSerArgGlnGly. 183
571CGAGCCAGCGT.....GGGGACCACCTCTTCCCGGAGCAAGGCC 610

184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
611 ACGCCCTGGGGGCACTTCTCAGGCCAGACAGATTGATTCOCGCTGCGG 660

200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
661 ATCTCTGGTCCCAACCACTTTGTTGGTGCATCATCGGAAAGGAGGCTT 710

216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
711 GACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATA 760

233 rgLysGluAsnAlaGlyAlaGluLysSerIleThrIleLeuSerThr 249
761 GAAAAGAGAACTCTGGAGCTGCAGAGAAGCCTGTCCACCATCATGCCACC 810

250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
811 CCAGAGGGGACTTCTGAAGCATGCGGCATGATTCTTTGAAATCATGCAGAA 860

266 sGluAlaGlnAspIleLysPheThrGluLulIleProLeuLysIleLeuA 283
861 AGGCAGATGAGCAACAACTAGCGGAAGACATTCCTCTGAAATCTTGG 910

283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
911 CACACAATGGCTTGGTGGAAAGACTGATTGGAAGAGAGGAGGAGAAATTTG 960

300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGI 316
961 AAGAAAATTGAACATGAACAGGACCAAGATAACAATCTCATCTTTGCA 1010

316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
1011 GGATTGAGCATATACACCCGGAAGAACCATCACTGTGAAGGCGACAG 1060

333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
1061 TTGAGGCTGTGCCAGTCTGAGATAGAGATTATGAAGAAGCTGCGTGAG 1110

350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1111 GCCTTTGAAAATGATGCTGGCTGTTAAACCAACAAGCCAATCTGATCCC 1160

366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1161 AGGTTGAACTCAGCGCCTTGGCATCTTTTCAACAGGACTGTCCGTGC 1210

377ProProThrSer.....GlyMetProProProThr..... 386
1211 TATCTCCACAGCAGGCGCCCGCGGAGCTCCCGCGTCCCTTACCAC 1260

387SerGlyProProSerAlaMetThrPro..... 395
1261 CCCTTCACTACCCACTCCGGATCTTCTCCAGCCTGTACCCCCCATCACA 1310

396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1311 GTTTGGCCCGTTCCTCCGATCATCACTCTTATCCAGAGCAGGAGATTGCA 1360

409 isGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425
1361 ATCTCTTCATCCCAACCCAGGCTGTGGCGGCATCATCGGGAAGAGGGG 1410

426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1411 GCACACATCAACACAGCTGGGAGATTCCCGGAGCCTCTATCAAGATTGC 1460

442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1461 CCCTGCGGAAGGCCACAGCTCAGCGGAAGAGTGGTCTATCATCATCAGCGGC 1510

```
459 roProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1511 |CCCCGAAAGCCAGGTTCAAGGCCAGGGACGGAGTCTTTGGGAAACTGAAA 1560
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIle 492
1561 |GGGAAAGCTCTTAAACCCCAAGAAAGTGAAGCTGGAAGCGCATAT 1610
492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysT 509
1611 |CAGAGTGCCCTCTCCACAGCTGGCGGGTGATTGGCAAAAGTGGAAGA 1660
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArg 525
1661 |CGGTGAACAACTGCAGAACTTAAACAGTGCAGAACTCATCTGCTCGT 1710
526 AspGlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHi 542
1711 |GACCAACCCAGATGAAATGAGGAAGTGATCGTCAAGAAATTATCGGGCA 1760
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
1761 |CTTCCTTGCTACCCAGACTGCACAGCGCAAGATCAGGGAATTTGTACAAC 1810
559 InValLysGlnHisGlnGlnGln 566
1811 |AGGTGAAGCAGCAGGAGCAGAAA 1833
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-061-709-7

```
seq_documentation_block:
; Sequence 7, Application block
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tsung
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061.709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7
```

alignment_scores:
Quality: 1873.00 Length: 509
Ratio: 4.116 Gaps: 9
Percent Similarity: 89.391 Percent Identity: 74.656

alignment_block:

US-09-685-176 x US-09-061-709-7 ..

Align seg 1/1 to: US-09-061-709-7 from: 1 to: 1946

```
77 ArgGlnArgIleArgLysLeuGlnIleArgAsnIleProHisLeuG1 93
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 CGGGTTAGGAGCGGAAAAATCCAAATCCGAAATATTCACCCAGCTCCG 151
93 nTrpGluValLeuAspSerLeuLeuValGlnTyrGlyValValGluSerC 110
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 ATGGGAAGTACTGGACAGCTCTGCTGCTAGTATGTCAGTAGAAGT 201
```

```
110 ysGluGlnValAsnThrAspSerGluThrAlaValValAsnValThrTyr 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
202 GTGAGCAAGTGAACACCGAGAGTGAGACGCGAGTGTGAATGTCACTAT 251
127 SerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheG1 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 TCAACCGGGAGCAGACCGCAAGCATCATGAAGCTGAATGGCCACCA 301
143 nLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluMetAla 160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 GTTGGAGAACCATTGCCCTGAAGGCTCTTACATCCCGGATGAGCAGATAG 351
160 laGlnGlnAsnProLeuGlnGlnProArgGlyArgGlyLeuGlyGln 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
352 CA...CAGGGACCTGAGAATGGCGCGCAGGG.....GGCTTTGGCTCT 392
177 ArgGlySerSerArgGlnGlySerPro.....GlySerValSerLy 190
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 CGGGGTGACGCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCGACCAA 442
190 sGlnLysProCysAspLeuProLeuArgLeuValProThrGlnPheV 207
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 GCAGCAGCAAGTGGACATCCCCCTTCGGCTCTCGTGCCCGCAGCATG 492
207 aGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLys 223
493 TGGGTGCCATTATTGGCAAGAGGGGGCCACCATCCGCAACATCACAAA 542
224 GlnThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAla 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
543 CAGACCCAGTCCAAGATAGACGTGCATAGGAGGAGAACCGAGTGCAGC 592
240 aGluLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAla 257
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
593 TGAAGATGATCTTGGAGATTATGCATAAGAGAGGCTAAGGACACCAACG 692
257 ysLysSerIleLeuGluIleMethLysLysGluAlaGlnAspIleLysPhe 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 GTAAGATGATCTTGGAGATTATGCATAAGAGAGGCTAAGGACACCAACG 692
274 ThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyAr 290
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
693 GCTGACGAGGTTCCTCCCTGAAGATCCTGGCCCATTAATACTTTAGGGCG 742
290 gLeuIleGlyLysGluGlyArgAsnLeuLysIleGluGlnAspThrA 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
743 TCATTTGCCAAGAGAGGCGGAACCTGAAGAGGATGAGCAAGATACCG 792
307 spThrLysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnPro 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 AGACAAAATCACCATCTCCTCGTTGCAAGACCTTACCCCTTTACAAACCT 842
324 GluArgThrIleThrValLysGlyAsnValGluThrCysAlaLysAlaG1 340
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
843 GAGAGGACCATCATCTGTGAAGGGGGCCATCGAGATTGTTGAGGGCCGA 892
340 uGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlas 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
893 GCAGGAATAATAAGAAAGGTCGGGAGGCTATGAGAAATGATGTGGCTG 942
357 erMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeu 373
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
943 CCATGAGC.....TCTCACTGATCCCTGGCCTGAACCTGGCTGCTGTA 986
374 GlyLeuPheProThrSerGlyMetProProThrSerGlyProPr 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
987 GGTCTTTTCCAGCTTCATCCAGCGAGTCCCGC.....CCTCC 1027
390 oSerAlaMetThr.....ProProTyrProGlnPheGluGlnSer....G 404
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1028 CAGCAGGTTACTGGGGCTGCTCCCTATAGTCTCTTTATCGAGCTCCCG 1077
404 luThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaIle 420
```

```

1078 AGCAGGATGGTGCAGGTGTTTATCCCGCCAGGTGGCGCCATC 1127
421 lIeGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAl 437
1128 ATCGCAAGAGAGGGCAGCAGCATCAACAGCTCCCGGTTTGCAGGCG 1177
437 eSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMetV 454
1178 CTCCATCAAGATGTCACACCCGCAACACCTGACTCCAAAGTTCGTATGG 1227
454 alIleIleThrGlyProGluAlaGlnPheLysAlaGlnGlyArgIle 470
1228 TTATCATCACTGGACCGCCAGGCCCAATTCAAGGCTCAGGGAGAGATC 1277
471 TyrGlyLysIleLysGluGluAsnPheValSerProLysGluGluVally 487
1278 TATGCAAACTCAAGGAGGAGAACTTCTTGGTCCCAAGGAGGAAGTAA 1327
487 sLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleG 504
1328 GCTGGAGACCCACATACGTGTGCAGCATCAGCAGCTGGCCGGGTCAATG 1377
504 lYlysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGlu 520
1378 GCAAGGTGGAACACGGTGAACAGTTGCAGAAATTTGACGGCAGCTGAG 1427
521 ValValValProArgAspGlnThrProAspGluAsnAspGlnValValva 537
1428 GTGGTAGTACCAGAGACACGACCCCTGATGAGAACGACCAAGGTCATCGT 1477
537 lYsIleThrGlyHisPheTyrAlaCysGlnValAlaGlnArgLysIleG 554
1478 GAAATTCATCGGACATTTCTATGCGCATCAGATGGCTCAACGGAGATCC 1527
554 InGluIleLeuThrGlnValLysGlnHisGlnGlnLysAlaLeuGln 570
1528 GAGACATCCTGGCCCGCAGGTTAAGCAG...CAGCATCAGAGGGGA...CAG 1571
571 SerGlyProProGlnSerArgLys 579
1572 AGTAACCAAGGCCAGGCAGCGAGGAAG 1598

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-061-709-8

seq_documentation_block:

```

; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Kouth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-8

```

alignment_scores:
Quality: 1868.00 Length: 568
Ratio: 3.820 Gaps: 6

```

Percent Similarity: 86.092 Percent Identity: 64.789
alignment_block:
US-09-685-696-176 x US-09-061-709-8 ..
Align seg 1/1 to: US-09-061-709-8 from: 1 to: 3283

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaLaProSerAs 17
73 ATGAACAAGCTTTTACATCGGAACCTTGAGCCCGCGCTCACCGCGGACGA 122
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyPropheL 34
123 CCTCCGGCAGCTCTTTGGGGACAGGAAGCTGCCCTTGGCGGAGACAGTFC 172
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTyrAla 50
173 TGCTGAAGTCCGGCTACGCCCTTCGTGGACTACCCCGACCAAGACTGGGCC 222
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
223 ATCCGCGCCATCGAGACCCCTCTCGGCTAAAGTGAATTCATGGAAAAAT 272
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
273 CATGGAAGTTGATTACTCAGTCTCTAAAAAGCTAAGGAGCAGGAAAAATTC 322
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
323 AGATTCCGAACATCCCTCCTCCTCAGCTGAGTGGGAGGTGTGGATGGACTT 372
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
373 TTGGCTCAATATGGACAGTGGAGAATGTGGAACAAGTCAACACAGACAC 422
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
423 AGAAACCGCCGCTTGTCAACGCTCAGATATGCAACAAGAGAGAGAAAAA 472
134 InAlaLeuAspLysLysGlnPheGlnLeuGluAsnPheThrLeuLys 150
473 TAGCCATGGAGAGCTAAGCGGCGCATCAGTTTGAGAACTACTCTCTCAAG 522
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
523 ATTTCTTACATCCCGATGAAGAGGTGAGCTCCCTTCGCCCTCAG... 570
167 nProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGly. 183
571 ....CGAGCCCGAGCGT.....GGGGACCACTCTTCCCGGAGCAAGGCC 610
184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
611 ACGCCCTCGGGGCGACTTCTCAGCCAGACAGAGATTGATTTCCCGCTCGGG 660
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
661 ATCCTGGTCCCGCCACCGATTTGTTGGTGCCATCATCGGAAAGGAGGCTT 710
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
711 GACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATA 760
233 rGlyGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
761 GAAAGAGAACTCTGGAGCTGCAGAGAGAGCTGTACCATTCATGCCACC 810
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
811 CCAGAGGGGACTTCTGAAGCATGCCGATGATTTCTTCAAAATCATGCAGAA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283

```

```
861 AGAGGCGATGAGACCAAACTACCGGAAGAGATTCTCTGAAATCTTGG 910
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
|||||.....|.....|.....|.....|.....|.....|.....|
911 CACACATGGCTGGTGGAGACTGATGGAAAAAGAGCGCAAAATTGG 960
300 LysLysIleGluGlnAspThrAspThrLysLysIleThrIleSerProLeuG1 316
|||||.....|.....|.....|.....|.....|.....|.....|
961 AAGAAAAATGAACATGAACAGGAGGACCAAGATAAACAATCTCATCTTGCA 1010
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
|||||.....|.....|.....|.....|.....|.....|.....|
1011 GGATTTGACATATACACCCGGAAGAACCATCACTGTGAAGGGACAG 1060
333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
|||||.....|.....|.....|.....|.....|.....|.....|
1061 TTGAGGCTGTGCCGTGCTGAGATAGATTATGAAGAGCTGCGTGAG 1110
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
|||||.....|.....|.....|.....|.....|.....|.....|
1111 GCCTTTGAAATCATATGCTGGCTGTAAACACCCACTCC..... 1149
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetP 383
|||||.....|.....|.....|.....|.....|.....|.....|
1150 .....|.....|.....|.....|.....|.....|.....| 1158
383 roProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
|||||.....|.....|.....|.....|.....|.....|.....|
1159 .....TCCACCTGTACCCCCATCACCAGTTTGGCCCGTTCGCCGAT 1200
400 PheGluGln...SerGluThrGluThrValHisGlnPheIleProAlaLe 415
|||||.....|.....|.....|.....|.....|.....|.....|
1201 CATCACTCTTATCCAGAGGAGATTGTAATCTCTTCATCCCAACCCA 1250
415 uSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuS 432
|||||.....|.....|.....|.....|.....|.....|.....|
1251 GGCTGTGGGGCCCATCATCGGGAAGAGGGGGCACACATCAAAACAGCTGG 1300
432 erArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAsp 448
|||||.....|.....|.....|.....|.....|.....|.....|
1301 CGAGATTCGGCGAGGCTCTATCAAGATGCCCTCGGAGGCCCGACAG 1350
449 AlaLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLy 465
|||||.....|.....|.....|.....|.....|.....|.....|
1351 GTCAGCGAAAGGATGTCATCATCACCGGGCCACCGGAAGCCAGTTCAA 1400
465 sAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerP 482
|||||.....|.....|.....|.....|.....|.....|.....|
1401 GGCCCGAGGACGGATCTTTGGGAAACTGAAAGAGGAAACTTCTTTAACC 1450
482 roLysGluGluValLysLeuGluAlaHisIleArgValProSerPheAla 498
|||||.....|.....|.....|.....|.....|.....|.....|
1451 CCAAGAGAGAGTGAAGCTGGAAGCGCATATCAGATGCCCTCTTCACA 1500
499 AlaGlyArgValIleGlyLysGlyLysGlyLysThrValAsnGluLeuGlnAs 515
|||||.....|.....|.....|.....|.....|.....|.....|
1501 GCTGGCGGGTGATTGGCAAAGTGGAAGACCGTGAACCAACTGCAGAA 1550
515 nLeuSerSerAlaGluValValProArgAspGlnThrProAspGluA 532
|||||.....|.....|.....|.....|.....|.....|.....|
1551 CTTAACCAAGTGAAGATCATCTGCTGCTGACCAACCCAGATGAAA 1600
532 snAspGlnValValValLysIleThrGlyHisPheTyrAlaCysGlnVal 548
|||||.....|.....|.....|.....|.....|.....|.....|
1601 ATGAGGAAGTGAAGTGAAGTATTCGGGCACCTCTTTGCTACGCCAGACT 1650
549 AlaGlnArgLysIleGlnGluIleLeuThrValLysGlnHisGlnG1 565
|||||.....|.....|.....|.....|.....|.....|.....|
1651 GCACAGCGCAAGATCAGGGAATTTGTACACACAGGTGAGGAGGAGCA 1700
565 nGln 566
|...
1701 GAAA 1704
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-061-709-5

seq_documentation_block:

; Sequence 5, Application US/09061709B
; Patent No. 6297364

; GENERAL INFORMATION:

; APPLICANT: Chen, Yao-tseng

; APPLICANT: Gure, Ali

; APPLICANT: Tsang, Solam

; APPLICANT: Stockert, Elisabeth

; APPLICANT: Jager, Elke

; APPLICANT: Knuth, Alexander

; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

; FILE REFERENCE: LUD 5538

; CURRENT APPLICATION NUMBER: US/09/061,709B

; CURRENT FILING DATE: 1998-04-17

; NUMBER OF SEQ ID NOS: 8

; SEQ ID NO 5

; LENGTH: 1708

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

US-09-061-709-5

alignment_scores:

Quality: 1637.00 Length: 453

Ratio: 4.093 Gaps: 9

Percent Similarity: 88.300 Percent Identity: 74.172

alignment_block:

US-09-685-696-176 x US-09-061-709-5 ..

Align seg 1/1 to: US-09-061-709-5 from: 1 to: 1708

133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLe 149

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

32 CGGGGAGCCATCATGAAGCTGAATGGCCACCATGTGGAGAACCATGCCCT 81

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

149 uLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuG 166

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

82 GAAGGTCCTCTACATCCCCGATGAGCAGATAGCA...CAGGGACCTGAGA 128

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

166 lnGlnProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGln 182

:: ||||| ||||| ||||| ||||| ||||| ||||| |||||

129 ATGGGCGCCGAGGG.....GGCTTTTGCTCTCGGGGTGAGCCCGCCAG 172

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

183 GlySerPro.....GlySerValSerLysGlnLysProCysAspLe 196

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

173 GGCTCACCTGTGGCAGCGGGGGCCCCAGCCAGCAGCAGCAAGTGGACAT 222

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

196 uProLeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyL 213

:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

223 CCCCCTCGGCTCTGTGTGCCACCCAGTATGTGGTGCCATTATTGGCA 272

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

213 ysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIle 229

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

273 AGGAGGGGGCCACCATCCGCAACATCAAAACAGACCCAGTCCCAAGATA 322

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

230 AspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIl 246

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

323 GACGTGATAGGAAGGAGACGCGAGGTGCAGCTGAAAGAGCCATCAGTGT 372

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

246 eLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluI 263

:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

373 GCACCTCACCCCTGAGGGCTGCTCCTCGGCTTGTAAAGATGATCTTGGAGA 422

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

263 leMetHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu 279

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

423 TTATGCATAAAGAGGCTAAGGACACCAACAGCGGCTGACGAGGTTCCTCG 472

```
280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlu1 296
|||||
473 AAGATCCTGGCCCAATAAATACTTTAGGGCTCTCATTTGGCAAGGAAG 522
|||||
296 YArgAsnLeuLysIleGluAspThrAspThrLysIleThrIleS 313
|||||
523 ACGGAACCTGAAGAGGAGTAGAGCAAGATACCGAGACAAAATCACCATCT 572
|||||
313 erProLeuGlnGluLeuThrLeuTyAsnProGluArgThrIleThrVal 329
|||||
573 CCTCGTTTCAAGACCTTACCCTTTACAAACCTTGAGAGGACCATCACTGTG 622
|||||
330 LysGlyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLy 346
|||||
623 AAGGGGCGCATCGAATTTGTTCAGGGCCGAGCAGGAAATAATGAAGAA 672
|||||
346 sIleArgGluSerTyArgGluAsnAspIleAlaSerMetAsnLeuGlnAlaH 363
|||||
673 AGTTGGGAGGCGCTATGAGATGATGTGCTGCCATGAGC.....TCTC 716
|||||
363 IsLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThr 379
|||||
717 ACCTGATCCTGGCTGAACCTGGCTGCTGTAGTCTTTTCCACGCTTCA 766
|||||
380 SerGlyMetProProThrSerGlyProProSerAlaMetThr..... 394
|||||
767 TCCAGCGCAGTCCGCGCC.....CCTCCAGCAGCGTTACTGGGGC 807
|||||
395 .ProProTyProGlnPheGluGlnSer...GluThrGluThrValHisG 410
|||||
808 TGCTCCCTATAGCTCTTTATGCAAGTTCCTCCGAGCAGGAGATGTGCAG 857
|||||
410 InPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGln 426
|||||
858 TGTATTATCCCGCCCGCAGCATGGCGCCATCATCGGCAAGAGGGCGAG 907
|||||
427 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaPr 443
|||||
908 CACATCAACAGCTCTCCGGTTTGGCCAGCGCTCCCATCAAGATTGCACC 957
|||||
443 oAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyProp 460
|||||
958 ACCGAAACACCTGACTGCCAAGTTCGTATGTTATCATCATCGGACCGC 1007
|||||
460 roGluAlaGlnPheLysAlaGlnGlyArgIleTyArgLysIleLysGlu 476
|||||
1008 CAGAGGCCCAATTCAAGGCTCAGGGAAGATCTATGGCAAACTCAAGGAG 1057
|||||
477 GluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleAr 493
|||||
1058 GAGAACTTCTTTGGTCCCAAGGGAAGTGAAGCTGGAGACCCACATACG 1107
|||||
493 gValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrV 510
|||||
1108 TGTCCAGCATCAGCAGCTGGCCGGGTTCATTGGCAAAAGTGGAAAAACGG 1157
|||||
510 aLysGluLeuGlnAsnLeuSerAlaGluValValValProArgAsp 526
|||||
1158 TGAACGAGTTCAGCAATTTGACGCGCAGCTGAGGTGGTAGTACCAAGAGAC 1207
|||||
527 GlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPh 543
|||||
1208 CAGACCCCTGTATGAGACGACGACGATCATCGTGAAATCATCGGACATTT 1257
|||||
543 eTyAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnV 560
|||||
1258 CTATGCCAGTCATGAGTGCCTAACCGAGATCCGAGACATCTCTGGCCGAGG 1307
|||||
560 aLysGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSer 576
|||||
1308 TTAAGCAG...CAGCATCAGAAGGA...CAGAGTAACCGGCCGAGCA 1351
|||||
```

```
577 ArgArgLys 579
|||||
1352 CGAGGAAG 1360

seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-08-021-608D-1

seq_documentation_block:
; Sequence 1, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2384
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 473 bp
; OTHER INFORMATION: variable region where R is A or G.
US-08-021-608D-1

alignment_scores:
Quality: 241.00 Length: 449
Ratio: 1.076 Gaps: 17
Percent Similarity: 49.889 Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-021-608D-1 ..
Align seg 1/1 to: US-08-021-608D-1 from: 1 to: 2384
161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
|||||
213 CAAAAAGACCTTTAGAGATGGAGATCAACCATGCTTAAGAAAGTGC 262
|||||
```



```
174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
263 TCCTCAAAATGACTCTTTTGGACACAGATTACCCAGCATCATCAGCAGC 312
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
191 lnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
313 AAAGCAGATCTGTAATGACAGAAGATAACAAGTTCAGATGGAATGGTT 362
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
208 GlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
363 GGATTCAATATTGGACAGAGGTGAACAGATCTCAGCATACAACAGGA 412
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAla 241
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
413 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGGTGGCTTCCAG 459
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
241 lLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
460 AAAGGTCCTGTATTTAACTGGAAACACCTGAATCTGTCCAGTCAGCAAAA 509
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
510 CGGTTACTGACACAGATTGTTGAAAAGGAAGACACCTCCTGGCTTCCA 559
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
560 TCATGGCGATGGACCGGAATACAGTTCAGAAATCATGATTCACGCTA 609
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
610 GCAAGGCAGGATTAGTCAATTTGGAAGGGGAGAACTATTAAACAGCTT 659
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
660 CAGAACGGGCTGGAGTAAATAGGTTATG.....ATTCAAGAC..... 698
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
319 rLeuTyrAsnProGluArgThr.....IleThrValLysG 331
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
699 .....GGGCCGACAGAACACTGGTGTGACAAACCTCTTAGGATTACAG 741
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
331 lYasnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
742 GAGACCCATATAAAGTTCAACAGCCCAAGGAAATGGTGTAGATTAAAT 791
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetas 359
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
792 CGTGATCAAGCGGTTTCAGAGAAGTTCGGAATGAGTATGGTCAAGA.. 839
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaGlyLeuP 376
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
840 .....ATAGGAGGA..... 848
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
376 heProThrSerGlyMetProProProThrSerGlyProProSerAla 392
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
849 .....AATGAAGGATAGATGTCGCC..... 869
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
393 MetThrProProThrProGlnPheGluGlnSerGluThrGluThrValHi 409
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
869 ..... 869
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 426
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
870 .....ATCCAGAGATTCTGTGTCATGTAATAGGAGAAATCGAG 912
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
426 lnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
913 AGATGATCAAAAAAATACAAAATGATCTGGTGTTCGCAATTCAGTTAAG 962
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
963 CCAGATGATGGCAACACCCGAA...AGGATAGCACAAAATAACAGGACC 1009
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
```

```

; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 473 bp
; OTHER INFORMATION: variable region where R is A or G.
US-08-726-160-1

alignment_scores:
  Quality: 241.00      Length: 449
  Ratio: 1.076        Gaps: 17
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-726-160-1 ..
Align seg 1/1 to: US-08-726-160-1 from: 1 to: 2384

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgArgGlyLe 174
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 CAAAAAGACCTTTAGAGATGGAGATCAACACAGATGCTAAGAAAGTTGC 262
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
174 uGlyGlnArgGlySerSerArgGlnGlnGlySerProGlySerValSerLysG 191
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
263 TCCTCAAAATGACTCTTTGTTGAAACACAGTTTACCACCGCATGCAGCAGC 312
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
191 InLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
313 AAACGAGATCTGTATGACAGCAAGATCAAAAGTTCCAGATGGATGGTT 362
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 GlyAlaIleLeuGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
363 GGATTCAATATGGCAGAGGAGTGAACAGATCTCAGCATACACAGGA 412
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaLag 241
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
413 ATCTGGATGCAAAATACAGATA...GCTCTGACAGTGGTGGCTTCCAG 459
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
241 LuLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
460 AAGGTCCTGTATTTAACTGGAACACCTGATCTGTCACGTCAGCAAAA 509
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
510 CGGTTACTGGACCAAGATTGTTGAAAAAGGAAGACAGCTCCTGCTTCCA 559
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
560 TCATGGCGGACGCGGAAATCAGTTCAAGAAATCATGATTCACGCTA 609
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
610 GCAGGCAGGATTAGTCATTGGAAGAGGGGAGAACTATTAAACAGCTT 659
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
660 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 698
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
319 rLeuTyAsnProGluArgThr.....IleThrValLysG 331
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
699 .....GGGCCGACAGACACTGTGTGTCAGCAAAACCTCTTAGGATTACAG 741
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
742 GAGACCCATATAAAGTTCAACACCCAGGAAATGGTTAGATTAAATT 791
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```

348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
792 CGTGATCAAGCGGTTTCAGAGAGTTCCGAATGAGTATGGTCAAGA... 839
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyLeuP 376
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
840 .....ATAGGAGGA..... 848
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
376 heProProThrSerGlyMetProProThrSerGlyProProSerAla 392
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
849 .....AATGAGGGATAGATGTCGCC..... 869
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
393 MetThrProProTyrProGlnPheGluInSerGluThrGluThrValHi 409
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
869 ..... 869
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 .....ATCCAGATTGCTGTTGGCATTGTAATAGGAAGAAATGGAG 912
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
426 InHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
913 AGATCATCAAAAAATACAAAATGATGCTGGTTCGCATTTCAGTTTAAAG 962
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
443 ProLaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
963 CCAGATGATGGACAAACACCCGAA...AGGATAGCACAATAACAGGACC 1009
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
459 oProGlu...AlaGlnPheLysAlaGln..... 467
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1010 TCCAGACCGATGTCAACATGCTGCAGAAATATTATACAGACCTTCTTCGAA 1059
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
468 .....GlyArgIleTyr 471
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1060 GTGTTTCAGCTGGTAATCCTGCTGGACCTGGACCTGGTGGTCGAGGAAGA 1109
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 GlyLys.....IleLysG 476
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1110 GGTAGAGGTCAAGGCAACTGGAACATGGGACCACCTGGTGGATTACAGGA 1159
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
476 uGluAsnPheValSerProLysGluGluValLysLeuAlaHisIleA 493
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1160 ATTAATTTTATT..... 1172
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1173 ..GTGCCAACTGGGAAACTGGATTATAATAGGAAAGAGGTCAAAACC 1220
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValValProArgAs 526
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1221 ATAAAAGCATAGCCAGCAGCTCTGGTGCAGAAATAGAACTTCAGAGAAA 1270
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1271 TCCTCCCAAAATGCAGATCCTAATATGAAGTTATTACAAATTCGTGGCA 1320
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
542 is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1371 CTCACACACATAGACTATGCTCGCAACTCATAGAGAAAAGATT 1367
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cgn2_5/ptodata/2/ina/PCUS_COMB.seq:PCT-US94-01782-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
```

```

; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2384
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE: HL60
; ORGANELLE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 473 bp
; OTHER INFORMATION: variable region where R is A or G.
PCT-US94-01782-1

```

```

alignment_scores:
  Quality: 241.00      Length: 449
  Ratio: 1.076        Gaps: 17
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x PCT-US94-01782-1 ..
Align seg 1/1 to: PCT-US94-01782-1 from: 1 to: 2384

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
|||||.....|
213 CAAAAAGACCTTTAGAGATGGAGATCAACCGATGCTTAAGAAAGTTGC 262
|||||.....|
174 uGlyGlnArgGlySerArgGlnGlySerProGlySerValSerLysG 191
|||||.....|
263 TCCTCAAAATGACTCTTTTGGAAACAGATTACCACCGATGCATCAGCAGC 312

```

```

191 InLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
|||||.....|
313 AAAGCAGATCTGTAAATGACAGAAGATAACAAGTTCCAGATGGAATGGTT 362

208 GlyAlaIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
|||||.....|
363 GGATTCATAATGGCAGAGGAGGTGAACAGATCTCAGCATACACAGGA 412

224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaG 241
|||||.....|
413 ATCTGGATGCAAAATACAGATA...GCTCTGACAGTGGTGGCTTCCAG 459

241 LuLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
|||||.....|
460 AAAGTCTCTGTATRTTAACCTGGAAACACCTGCAATCTCTCCAGTCAGCAAAA 509

258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPhe 274
|||||.....|
510 CGGTTACTGGACACAGATTGTTGAAAAAGGAAGACACAGCTCCTGGCTTCCA 559

274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
|||||.....|
560 TCATGGCGATGGACCGGAAATGCAGTTCAAGAAATCATGATTCCAGCTA 609

286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIle 302
|||||.....|
610 GCAAGGCAGGATTAGTCATTGGAAAAGGGGAGAACTATTAAACAGCTT 659

303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeu 319
|||||.....|
660 CAGGAACGGGTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 698

319 rLeuTyrAsnProGluArgThr.....IleThrValLysG 331
|||||.....|
699 .....GGGCCGACAGACACTGGTCTGCACAAACCTCTTAGGATTACAG 741

331 LyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||.....|
742 GAGACCCATATAAAGTTCAACAAGCCCAAGGAATGGTTAGAGTTAATT 791

348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
|||||.....|
792 CGTGATCAAGCGGTTTCAGAGAAGTTCCGAATGAGTATGGGTCAAGA... 839

359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeu 376
|||||.....|
840 .....ATAGGAGGA..... 848

376 heProProThrSerGlyMetProProThrSerGlyProProSerAla 392
|||||.....|
849 .....AATGAAGGGATAGATGTCCTCC..... 869

393 MetThrProProTyrProGlnPheGluGlnSerGluThrGluThrValH 409
|||||.....|
869 ..... 869

409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 426
|||||.....|
870 .....ATTCCAAGATTTGCTTGGCATTGTAATAGGAAGAAATGGAG 912

426 InHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
|||||.....|
913 AGATGATCAAAAAAATACAAATGATGCTGGTGTTCGCATTCAGTTTAAG 962

443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
|||||.....|
963 CCAGATGATGGGACACACCCGAA...AGGATAGCAAAATAACAGGACC 1009

459 oProGlu...AlaGlnPheLysAlaGln..... 467
|||||.....|
1010 TCACAGACCGATGTCACATGCTGCAGAAATATTATACAGACCTTCTTCGAA 1059

```

```
468 .....GlyArgIleTyr 471
1060 GTGTTACGGCTGGTAATCCTGGTGGACCTGGACCTGGTGGTCGAGAGA 1109
472 GlyLys .....IleLysG1 476
1110 GGTAGAGTCAAGGCACACTGGAACATGGGACCACCTGGTGATTACAGGA 1159
476 uGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleA 493
1160 ATTTAATTTATT ..... 1172
493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
1173 .GTGCCAACTGGGAAACTGGATTAATAATAGGAAAGGAGGTGAAC 1220
510 ValAsnGluLeuGluAsnLeuSerSerAlaGluValValProArgAs 526
1221 ATAAAAGCATNAGCCAGCTCTGGTGCAGAAATAGAACTTCAGAGAAA 1270
526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
1271 TCCTCCACCAATGCAGATCTAATATAGAGTTATTACAAATTCGTGCA 1320
542 is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
1321 CTCACACACAGATAGACTATGCTCGCAACTCATAGAAGAAAGATT 1367
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-021-608D-9
seq_documentation_block:
; Sequence 9, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: CDNA
```

```
; HYPOTHETICAL: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 470 bp variable
; OTHER INFORMATION: region where R is A or G.
US-08-021-608D-9

alignment_scores:
  Quality: 236.50      Length: 449
  Ratio: 1.056        Gaps: 18
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-021-608D-9 ..
Align seg 1/1 to: US-08-021-608D-9 from: 1 to: 2381
161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
263 TCCTCAAAATGACTCTTTTGGACACACAGTTACCA...CCGATGCATCAGC 309
191 lnLysProCysAspLeuProLeuArgLeuValProThrGlnPheVal 207
310 AGCAAAGATCTGTAATGACAGAAGAATAACAAAGTTCCAGATGGAATGGTT 359
208 GlyAlaIleIleGlyLysGluGluAlaThrIleArgAsnIleThrLysG1 224
360 GGATTCAATATGGCAGGAGTGACACAGATCTCACGCATACACACAGA 409
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGluAlaLag 241
410 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGGTGGCTTCCAG 456
241 luLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
457 AAAGTCTCTGTATTTAACTGGAACACCTGAATCTGTCCAGTCAGCAAAA 506
258 LysSerIleLeuGluIleMetHisLysGluAlaGlnAspIleLysPheTh 274
507 CGGTACTGGACAGATTGTTGAAAAGGAGACACAGCTCTCTGGCTTCCA 556
274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsnA 286
557 TCATGGCGATGGACCGGAAATGCAGTTCAAGAAATCATGATTCAGCTA 606
286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
607 GCAAGGCAGGATTAGTTCATTTGAAAAGGGGAGAACTATTAAACAGCTT 656
303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
657 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAGAC..... 695
319 rLeuTyrAsnProGluArgThr.....IleThrValLysG 331
696 .....GGGCCGAGAACACTGGTGTGCACAAACCTCTTAGGATTACAG 738
331 lyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIle 347
739 GAGACCCATATAAAGTTCAACAGCCAGGAAATGGTGTAGAGTTAATT 788
348 ArgGlu.....SerTyrGluAsnAspIleAlaSerMetAs 359
789 CGTGATCAAGCGGCTTTTCAGAGAAGTTCCGGAAATGAGTATGGTCAAGA... 836
```

```
359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuP 376
    ||| |||
837 .....ATAGGAGGA..... 845
376 heProProThrSerGlyMetProProThrSerGlyProProSerAla 392
    : : : : :
846 .....AATGAAGGATAGATGTCCTCC..... 866
393 MetThrProProTyrProGlnPheGluGlnSerGluThrGluThrValHI 409
    : : : : :
866 ..... 866
409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyC 426
    ||||| : : : : :
867 .....ATCCAAAGATTGCTGTGGCAATTGTAATAGGAAGAAATGGAG 909
    : : : : :
426 InHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
    : : ||||| : : : : :
910 AGATGATCAAAAAAATACAAAATGATGCTGGTTCGCATTCAGTTTAAG 959
    : : : : :
443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
    ||| : : : : :
960 CCAGATGATGGGCAACACCCGAA...AGGATAGCACAATAACAGGACC 1006
    : : : : :
459 oProGlu...AlaGlnPheLysIleGln..... 467
    ||||| : : : : :
1007 TCCAGACCGATGTCAACATGCTGCAGAAATATTACAGACCTTCTTCGAA 1056
    : : : : :
468 .....GlyArgIleTyr 471
1057 GTCTTCAGCTGTATCTCTGGTGGACCTGGTGGTTCGAGGAAGA 1106
    : : : : :
472 GlyLys.....IleLysGI 476
    ||| : : : : :
1107 GGTAGAGTCAAGCACTGGAACATGGGACCACCTGGTGGATTACAGGA 1156
    : : : : :
476 uGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleA 493
    | ||||| : : : : :
1157 ATTTAATTTTATT..... 1169
493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
    ||||| : : : : :
1170 ..GTGCCAACTGGGAAACTGGATTATTAATAGGAAAGAGGTGAACC 1217
    : : : : :
510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArgAs 526
    : : : : :
1218 ATAAAGCATAGCAGCAGCTGTGGCAAGATAGAACTTCAGAGAAA 1267
    : : : : :
526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
    : ||||| : : : : :
1268 TCCTCCACCAATGCAGATCCTAATATGAAGTTATTACAAATTCGTGGCA 1317
    : : : : :
542 is.....PheTyrAlaCysGlnValAlaGlnArgLysIle 553
    ||||| ||||| : : : : :
1318 CTCACACAGATAGACTATGCTCGGCACTCATAGAAAGAAAGATT 1364
    : : : : :
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-726-160-9

seq_documentation_block:
; Sequence 9, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVITAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
```

```

; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION: 470 bp variable
; OTHER INFORMATION: region where R is A or G.
US-08-726-160-9

alignment_scores:
    Quality: 236.50      Length: 449
    Ratio: 1.056        Gaps: 18
    Percent Similarity: 49.889    Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x US-08-726-160-9 ..
Align seg 1/1 to: US-08-726-160-9 from: 1 to: 2381

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgArgGlyLe 174
    ||| : : : : :
213 CAAAAAAGACCTTTAGAAGATGGAGATCAACACAGATGCTAAGAAAGTTGC 262
    : : : : :
174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
    ||||| : : : : :
263 TCCTCAAAATGACTCTTTTGGACACACAGATTACCA...CCGATGCATCAGC 309
    : : : : :
191 InLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheVal 207
    ||||| : : : : :
310 AGCAAGATCTGTATACAGAGAGATACAAAGTTCAGATGGAATGGTT 359
    : : : : :
208 GlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGI 224
    ||| ||||| : : : : :
360 GGATTCATAATTGCACAGAGGAGTGAACAGATCTCAGCATACACAGGA 409
    : : : : :
224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaIaG 241
    : : : : :
410 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGGTGCCTTCAG 456
    : : : : :
241 LuLysSerIleThrIleLeuSerThrProGluGlyThrSerAlaAlaCys 257
    ||||| : : : : :
457 AAAGGTCTCTGTATTTAACTGGAACACCTGAATCTGCTCCAGTCAGCAAA 506
```

```

1218 ATAAAAGCATAGCCAGCAGCTCTGTCGCAAGAATAAGAACTTCAGAGAAA 1267
      ::::::::::::::: |||::|||::: :::: |||::
526 pcInThrProAspGluAsnAspGlnVal...ValValLysIleThrClyH 542
      : |||::: ::::: ::::: ::::: ||| |||
1268 TCCTCACCAAAATGACGATCCCTAATATGAAGTTATTTACAAATTCGGGCA 1317
      ||| ||| |||::: ::::: ||| |||
542 is.....PheTyraIaCysGlnValaIaGlnAArgLysIle 553
      ||||| |||::: ::::: ||| |||
1318 CTCACACACAGATAGACTATGTCGGCCACTCATAGAAGAAAAGATT 1364
      ::::::::::::::: |||::|||::: :::: |||::

seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:pCT-US94-01782-9
seq_documentation_block:
; Sequence 9, Application:
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: cDNA
; HYPOTHEetical: No
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE: HL60
; ORGANELLE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:

```

```

; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 470 bp variable
; OTHER INFORMATION: region where R is A or G.
PCT-US94-01782-9

alignment_scores:
  Quality: 236.50      Length: 449
  Ratio: 1.056         Gaps: 18
  Percent Similarity: 49.889  Percent Identity: 22.272

alignment_block:
US-09-685-696-176 x PCT-US94-01782-9
Align seg 1/1 to: PCT-US94-01782-9 from: 1 to: 2381

161 GlnGlnAsnProLeuGln.....GlnProArgGlyArgGlyLe 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 CAAAAAGACCTTTAGAGATGGAGATCAACCCAGATGCTAAGAAAGTTGC 262

174 uGlyGlnArgGlySerSerArgGlnGlySerProGlySerValSerLysG 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 TCCTCAAAATGACTCTTTGGAAACAGTTACCA...CGATGCGATCAGC 309

191 lnLysProCysAspLeuProLeuArgLeuValProThrGlnPheVal 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 AGCAAGATCTGTATGACAGAGAGATACAAAGTTCCAGATGGATGGTT 359

208 GlyAlaIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysG 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
360 GGATTCAATAATGGCAGAGAGGTGAACAGATCTCAGCGCATACAACAGGA 409

224 nThrGlnSerLysIleAspValHisArgLysGluAsnAlaGlyAlaAag 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 ATCTGGATGCAAAATACAGATA...GCTCCTGACAGTGTGGCCTTCCAG 456

241 luLysSerIleThrLeuSerThrProGluGlyThrSerAlaAlaCys 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 AAAGTCCTGTATRTAATCTGGACACCTGATCTGTCAGTCAGCAAAA 506

258 LysSerIleLeuGluIleMethHisLysGluAlaGlnAspIleLysPheTh 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
507 CGGTTACTGGACAGATTTGTAAGAAAGGAAGACAGCTCTCTGGCTTCCA 556

274 rGlu.....GluIleProLeuLysIleLeuAlaHisAsna 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
557 TCATGGCGATGGACCGGAAATGCGAGTTCAAGAAATCATGATTCACGCTA 606

286 snPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIle 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
607 GCAAGCAGGATGATGTCATGGAAGAGGGGAGAGAACTATTAAACAGCTT 656

303 GluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuTh 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 CAGGAACGGCTGGAGTTAAATGGTTATG.....ATTCAAGAC..... 695

319 rLeuTyAsnProGluArgThr.....IleThrValLysG 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
696 .....GGGCGCGAGAACACTGGTCTGCACAAACCTCTTAGGATTACAG 738

331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
739 GAGACCCATATAAGTTCAACAGCAAGCAAGAAATGGTGTAGATTAAAT 788

348 ArgGlu.....SerTyGluAsnAspIleAlaSerMetAs 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
789 CGTGATCAAGCGGTTTCAGAGAAAGTTCGGAATGAGTATGGGTCAAGA.. 836

359 nLeuGlnAlaHisLeuIleProGlyLeuAsnLeuAlaLeuGlyLeup 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
837 .....ATAGGAGGA..... 845

```

```

376 heProProThrSerGlyMetProProProThrSerGlyProProSerAla 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
846 .....AATGAAGGATAGATGTCCCC..... 866

393 MetThrProProTyProGlnPheGluGlnSerGluThrGluThrValHi 409
866 ..... 866

409 sGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyG 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
867 .....ATTCCAAGATTTGCTGCTTGCATTGTAATAGGAGAAATGGAG 909

426 lnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAla 442
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
910 AGATGATCAAAAAAATACAAAATGATGCTGTGCTGCATTGAGTTAAG 959

443 ProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyPr 459
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
960 CCAGATGATGGGACAAACCCGAA...AGGATAGCACAAATAACAGGACC 1006

459 oProGlu...AlaGlnPheLysAlaGln..... 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1007 TCAGACCGATGTCAACATGCTCGAGAAATATTACAGACCTCTTCGAA 1056

468 .....GlyArgIleTy 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1057 GTCTTCAGGCTGTAACTCTGTGTGGACCTGGACCTGTGTGTCGAGGA 1106

472 GlyLys.....lleLysG 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1107 GGTAGAGGTCAAGGCAACTGGAACATGGGACCACTGGTGGATTACAGGA 1156

476 uGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisIleA 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1157 ATTTAATTTATT..... 1169

493 rgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThr 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1170 ..GTGCCAACTGGGAAACTGGATTAAATATAGAAAGGAGGTGAAACC 1217

510 ValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArgAs 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1218 ATAAAAGCATAAAGCCAGCAGTCTGTGTGCAAGAAATAGAACTTCAGAGAA 1267

526 pGlnThrProAspGluAsnAspGlnVal...ValValLysIleThrGlyH 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1268 TCTCCACCAATGCGAGATCTCTAATGAGTTATTACAAATTCGTGGCA 1317

542 is.....PheTyAlaCysGlnValAlaGlnArgLysIle 553
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1318 CTCACAACAGATAGATATGCTCGCAACTCATAGAGAAAGATT 1364

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-021-608D-7
seq_documentation_block:
; Sequence 7, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

```

```

/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WORDPERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/021,608D
/ FILING DATE: 22-FEB-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: WILLIAM S. FEILER
/ REGISTRATION NUMBER: 26,728
/ REFERENCE/DOCKET NUMBER: 2026-4063
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1803
/ TYPE: Nucleic acid
/ STRANDEDNESS: Double
/ TOPOLOGY: Unknown
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Human
/ CELL LINE: HL60
/ US-08-021-608D-7

```

```

alignment_scores:
  Quality: 229.50      Length: 405
  Ratio: 1.136         Gaps: 16
  Percent Similarity: 49.877      Percent Identity: 22.469

```

```

alignment_block:
  US-09-685-696-176 x US-08-021-608D-7

```

```

Align seg 1/1 to: US-08-021-608D-7 from: 1 to: 1803

```

```

202 ValProThrGlnPheValGlyAlaIleGlyLysGluGlyAlaThrII 218
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
295 GTCCAGATGAATGGTGGATTCAATATGGCAGGAGGTGAACAGAT 344
218 eArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgLysG 235
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
345 CTCACGCATACACAGGAATCTGGATGCAAAATACAGATA...GCTCTG 391
235 luAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerProGlu 251
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
392 ACAGTGTGGCTTCCAGAAAGTCTGTATGTTAACTGGGAACACCTGAA 441
252 GlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAl 268
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
442 TCTCTCAGTCAGCAAAACGGTACTGGACAGATTGTTGAAAAGGAAG 491
268 aGlnAspIleLysPheThrGlu.....GluIleProLeuL 280
||| :||| :||| :||| :||| :||| :||| :||| :|||
492 ACCAGCTCTGGCTTCCATCATGCGGATGACCGGGAATGCAGTTCAAG 541
280 ysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGly 296
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
542 AAATCATGATTCCAGCTAGCAGGAGGAGGATAGTCATTGGAAAAGGGGA 591
297 ArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSe 313
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
592 GAACCTATTAAACAGCTTCAGAACGGGCTGGAGTTAAATGGTTATG.. 639
313 rProLeuGlnGluLeuThrLeuThrAsnProGluArgThr..... 326
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
640 ....ATTCAAGAC.....GGGCGCAGCAACACTGGTGCTGACA. 673

```

```

327 .....IleThrValLysGlyAsnValGluThrCysAlaLysAlaGluGlu 341
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
674 AACCTCTTAGGATTACAGGAGACCCATATAAGTTCAACAGCAAGGAA 723
342 GluIleMetLysLysIleArgGlu.....SerTyrGluAs 353
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
724 ATGGTGTAGAGTTAAATTCGTGATCAAGCGGTTTCAGAGAAGTTCGGA 773
353 nAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsnL 370
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
774 TCAGTATGGGTCAGCA.....ATAGGAGCA..... 798
370 euAsnAlaLeuGlyLeuPheProThrSerGlyMetProProThr 386
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
799 .....AATGAAGGATAGATGTCCCC... 819
387 SerGlyProProSerAlaMetThrProTyrProGlnPheGluGlnSe 403
819 ..... 819
403 rGluThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAlaI 420
||||| :||| :||| :||| :||| :||| :||| :||| :|||
820 .....ATTCCAGATTTGCTGTGGCATG 844
420 lelleGlyLysGlnGlnHisIleLysGlnLeuSerArgPheAlaGly 436
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
845 TAATAGGAAGAAATGGAGAGATGATCAAAAAAATACAAAATGATGCTGT 894
437 AlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArgMe 453
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
895 GTTCGCATTCAGTTTAAGCCAGATGATGGGACACACCCGAA...AGGAT 941
453 tValIleIleThrGlyProGlu...AlaGlnPheLysAlaGln.... 467
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
942 AGCACAATAACAGGACCTCCAGACCGATGTCAACATCTCGAGAAATTA 991
467 ..... 467
992 TTACAGAGACCTTCTCGAAGTGTTCAGGCTGTGTAATCTGCTGGACCTGA 1041
468 .....GlyArgIleTyrGlyLys..... 473
||||| :||| :||| :||| :||| :||| :||| :||| :|||
1042 CCTGGTGTGCGAGGAAGAGGTAGAGGTCAAGGCAACTGGAACATGGGACC 1091
474 .....IleLysGluGluAsnPheValSerProLysGluGluValL 487
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1092 ACCTGGTGGATTACAGGAATTTAATTTATT..... 1122
487 ysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIle 503
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1123 .....GTGCCAACTGGGAAACCTGGATTATAATA 1152
504 GlyLysGlyGlyLysThrValAsnGlnLeuGlnAsnLeuSerSerAlaGl 520
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1153 GGAAAAGGAGGTGAACCATATAAAGCATAGCCAGCAGCTGTGGTCAAG 1202
520 uValValValProArgAspGlnThrProAspGluAsnAspGlnVal...V 536
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1203 AATAGAACTTCAGAGAATCTCCACCAATGTCAGATCCTTAATATGAAGT 1252
536 alValLysIleThrGlyHis.....PheTyrAlaCysGlnVal 548
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1253 TATTACAAATTCGTGGCAGCTCCACAACAGATAGACTATGTCTCGGCACTC 1302
549 AlaGlnArgLysIle 553
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
1303 ATAGAGAAAAGATT 1317

```

```

seq_name: /cqn_6/ptodata/2/ina/5A_COMB.seq:US-08-726-160-7

```

```

seq_documentation_block:
; Sequence 7, Application US/08726160

```


549 AlaGlnArgLysIle 553
 :::~::~|||||
1303 ATAGAAGAAAAGATT 1317

OM of: US-09-685-696-176 to: EST:* out_format : pfs

Date: Jan 25, 2002 10:46 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-Q/cgnt2/_uspto_spool/US09685696/runat_25012002.145550.10634/app_query.fasta_1.648
-DB=EST -OFT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCI=0.000 -LOAPEXT=0.000 -QGAPOP=4.000
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM-ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09685696 -CGN1_1.4152 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-685-696-176

Query length: 579

Database: EST.*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1537.690000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_hlc:AK011689	+	2858.00	4213.85	1.5e-225	2202	AK011689 Mus musculus 10 days
gb_hlc:AK013940	+	2192.00	3225.42	1.7e-170	2780	AK013940 Mus musculus 13 days
gb_est2:BF608004	+	1155.00	1702.09	1.2e-85	769	BF608004 MYL_000946 Mouse 9-day
gb_est1:AU127673	+	1065.50	1570.01	2.8e-78	733	AU127673 AU127673 NT2R2 Homo S
gb_est1:BE345335	+	1047.50	1542.64	9.2e-77	790	BE345335 601070391F1 NIH_MGC_12
gb_est1:AL584627	+	975.50	1437.09	7.0e-71	707	AL584627 AL584627 Stratagene CH
gb_est2:BG079106	+	952.00	1404.06	4.8e-69	587	BG079106 H3035G11-5 NIA Mouse 1
gb_est1:AA117282	+	936.50	1379.75	1.1e-67	677	AA117282 mn25a10.r1 Beddington
gb_est1:AA163596	+	911.00	1342.45	1.3e-65	645	AA163596 mn3802.r1 Beddington
gb_est1:AL584375	+	909.50	1340.20	1.7e-65	647	AL584375 AL584575 Stratagene CH
gb_est1:BF217988	+	903.00	1331.52	5.3e-65	586	BF217988 601862766F1 NIH_MGC_57
gb_est2:BF984367	+	890.00	1308.22	1.1e-63	896	BF984367 602308484F1 NIH_MGC_88
gb_est1:AU134323	+	884.00	1301.09	2.6e-63	745	AU134323 AU134323 OVARC1 Homo s
gb_est2:BG082113	+	868.00	1279.90	4.0e-62	573	BG082113 H3072F04-5 NIA Mouse 1
gb_est1:AA691146	+	835.50	1232.01	1.8e-59	559	AA691146 vt68009.s1 Knowles Sol
gb_est1:AL584570	+	831.50	1225.32	4.4e-59	606	AL584570 AL584570 Stratagene CH
gb_est2:BF217988	+	822.50	1210.91	2.8e-58	679	BF217988 601862766F1 NIH_MGC_57
gb_est1:AW342998	+	822.50	1210.89	2.8e-58	680	AW342998 fi172a06.y1 Sugano Kawa
gb_est2:BG748346	+	816.50	1199.46	1.2e-57	889	BG748346 602705902F1 NIH_MGC_43
gb_est2:BG069090	+	814.50	1198.57	1.3e-57	715	BG069090 H3072F04-3 NIA Mouse 1
gb_est2:BE869656	+	803.00	1179.33	1.6e-56	902	BE869656 601446879F1 NIH_MGC_65
gb_est1:AA349384	+	791.00	1167.30	7.4e-56	494	AA349384 vk86503.s1 Knowles Sol
gb_est1:AF046624	+	776.50	1142.22	1.9e-54	721	AF046624 AF046624 Mus musculus
gb_est1:AA116751	+	763.00	1125.00	1.7e-53	539	AA116751 mn21h01.r1 Beddington
gb_est1:AA587050	+	753.50	1109.60	2.7e-52	620	AA587050 AL587050 Stratagene CH
gb_est1:AA427803	+	735.00	1085.50	2.7e-51	439	AA427803 zw49c11.r1 Soares tota
gb_est1:AL558693	+	726.00	1089.53	2.1e-50	579	AL558693 fb79c12.y1 Zebrafish W
gb_est2:BG086000	+	721.50	1063.53	4.5e-50	540	BG086000 H3120E12-5 NIA Mouse 1
gb_est1:AL5866757	+	719.50	1059.32	7.7e-50	616	AL5866757 AL586757 Stratagene CH
gb_est1:AU134724	+	712.00	1047.08	3.4e-49	643	AU134724 AU134724 PLACE1 Homo s
gb_est1:AW343989	+	706.00	1038.08	1.2e-48	702	AW343989 fi172a06.x1 Sugano Kawa
gb_est2:BG107753	+	704.50	1029.50	3.5e-48	1368	BG107753 602277962F1 NIH_MGC_8
gb_est2:BG115319	+	684.00	1003.89	9.4e-47	831	BG115319 602336224F1 NIH_MGC_86
gb_est1:AU117945	+	676.00	991.24	4.8e-46	504	AU117945 AU117945 HEMBA1 Homo S
gb_est1:BE622021	+	663.50	974.40	4.1e-45	759	BE622021 601440673F1 NIH_MGC_72
gb_est2:BF185290	+	660.00	970.52	6.8e-45	662	BF185290 601843593F1 NIH_MGC_54
gb_est1:AA120364	+	658.00	969.04	8.2e-45	567	AA120364 mn17b02.r1 Beddington
gb_est1:AA141739	+	648.00	956.24	4.2e-44	459	AA141739 EST291818 Normalized
gb_est2:C77087	+	635.00	934.45	6.9e-43	599	C77087 C77087 Mouse 3.5-dpc bla
gb_est2:BE881835	+	633.50	928.25	1.5e-42	910	BE881835 601504962F1 NIH_MGC_71
gb_est2:BG499373	+	632.00	929.49	1.3e-42	633	BG499373 602547471F1 NIH_MGC_60

gb_est2:BF241103 + 631.00 924.38 2.5e-42 926 + BF241103 601880647F1 NIH_MGC
gb_est2:BF185289 + 620.50 912.15 1.2e-41 654 + BF185289 601843592F1 NIH_MGC
gb_est2:DS8801 + 620.00 916.11 7.3e-42 399 + DS8801 HUM512B06B Clontech h
gb_est1:AA073514 + 617.00 910.81 1.4e-41 437 + AA073514 mm83e03.r1 Stratage

seq_name: gb_hlc:AK011689

seq_documentation_block:

LOCUS AK011689 2202 bp mRNA HTc 05-JUL-2001
DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched
library, clone:2610036B18, full insert sequence.

ACCESSION AK011689

VERSION AK011689.1 GI:12847972

KEYWORDS CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library

clone:2610036B18.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2202)

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Methods in enzymology. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (bases 1 to 2202)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

JOURNAL 20499374

MEDLINE 11042159

REFERENCE 3 (bases 1 to 2202)

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multipillar sequencer

Genome research. 10 (11), 1757-1771 (2000)

JOURNAL 20530913

MEDLINE 11076861

REFERENCE 4 (bases 1 to 2202)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

JOURNAL 5 (bases 1 to 2202)

REFERENCE 5 (bases 1 to 2202)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hakagawa,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishii,K., Nomura,K., Numazaki,K., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for


```

/protein_id="BAB29071.1"
/db_xref="GI:12851514"
/translation="MNKLYIGNLNEVSVPADLEKVPFAEHKHSYSGQFLVKSGYAFVDC
PDEHMAKIEFTSGVLEQGRKLEIEHSVPKKORSKTOIRNIPPQLRWEVLDSLLA
QTGVWENCQVNTSESTAVVNTYSNREQTRQIMKLNGLHOLENHALKVSYIPDEQIT
OGPNRRGGFGSGOPROGSPVAGAPAKQOPVDIPLRLLVPTOVVGAIGKEGATI
RNITKOTOSKIDVHRKENAGAAKATSVHSTPEGCSSACKMILEIMHKEAKDTKADIE
VPLKILAHNNFVGRLLIGKEGRNLKQVEQDTEFKITLISSLDLTLNPERTITVVKAGIE
NCCRAQETMKKRYEYENDVAMSLSHLIFGLNLAAVGLFPASSAVVPPPPSSVTG
AAPTYSFQMAQPEQMFVIPAQAQGAITGKKGQHIKQLSRFASASIKIAPPETPDSK
VMVWITGQAFQAGRIYGLKEENFPKPEVKLETHIRVPSAAGRIVIGGGK
TVNELQNLTAEEVVPDQTPDENQVIVKIIGHFVASQMAQRKIRDIILAQVKQHQH
GOSNLAQARRK"
BASE COUNT      691 a      787 c      752 g      550 t
ORIGIN

```

```

alignment_scores:
  Quality: 2192.00      Length: 585
  Ratio: 4.152          Gaps: 8
  Percent Similarity: 90.256      Percent Identity: 74.188

```

alignment_block:

US-09-685-696-176 x AK013940

Align seg 1/1 to: AK013940 from: 1 to: 2780

```

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
279 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGTGTGACCCCGCAGA 328
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
329 CTTGGAGAAAGTATTCGGGAGACAAAGATCTCTACAGCGCGCAGTTCT 378
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
379 TGGTCAATCCGGCTACGCCCTTCGTGGATGTCGCCGACGACACTGGGCG 428
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
429 ATGAAGGCCATCGAAACTTTCTCGGGAAAGTAGAACTGCAAGGAAACG 478
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
479 TCTAGAGATTGAACACACTCAGTCCCAAAAACAAAGGAGTCCGGAATAAC 528
84 InileArgAsnIleProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
529 AGATTCCCATATTCACCTCAGCTCCGATGGGAAGTGCTAGATAGCCTG 578
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
579 CTGGCTCAGTACGCTACAGTGGGAACTGTGAGCAAGTGAACACTGAAG 628
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
629 TGAGACAGCTGTGTGTCACAGCTACTACTACTAACCGGAGCAGACACAGC 678
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
679 AAGTATCATGAAGCTAAATGGCCATCACTGGAGAACCATGCCCTGAAG 728
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
|||||
729 GTCCTCTACATACCTGATGACACATAACG...CAAGTCTCTGAGTAAG 775
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
|||||
776 GCGTCGTGGA.....GCGTTTGGGCTCGGGGCCACCCCGCAAGGGT 819
184 erPro.....GlySerValSerLysGlnLysProCysAspLeuPro 197
|||||

```

```

820 CGCCGCTGGCAGCAGGGGCTCCAGCCRAAGCAGCAGCCAGTGGACATCCCT 869
198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleLeuGlyLysGl 214
|||||
870 CTCGGGCTCTGTGTGCTTACGAGTATGTAGCGCTATCATTTGCAAGGA 919
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
|||||
920 GGGTGCACCACCTCGCAACATCACAAAACAGACAGCAGTCCAAAATAGACG 969
231 alHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
|||||
970 TGCATAGGAAGAGAAATCGGGCGCTCGGAGAAAGGCATCAGGTGCAT 1019
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMe 264
|||||
1020 TCAACCCCTGAAGGCTGCTCTCCGCGTGCAAGATGATCTTTGGAGATTAT 1069
264 thsLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
|||||
1070 GCACAAGGAGGCAAAAGGACACCAAAACGGCAGATGAAGTTCCTCGTAAGA 1119
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
|||||
1120 TCCTGGCTCATACAACATTCGTCGGGCGACTCATTTGGCAAGAAAGGGCGG 1169
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
|||||
1170 AACCTGAAGAGGTGGAGCAGACAGACAGAGCAAGATCACCATCTCATC 1219
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
|||||
1220 GCTCCAGGACCTCAGCTCTATAACCTCGAGGAGCAGCATCTGTGAAGG 1269
331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347
|||||
1270 CGCCATTTGAGAGCTGTTCAGGCGCCGAGGAGATCATGAAGAAAGTT 1319
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
|||||
1320 CGAGAGCTTTACGAGAACGAGCTGGCGCCATGAGTTCAGTCCCACCT 1369
364 uileProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
|||||
1370 CATCCCTGGGCTTAACTTGGCTGTGTAGGTCTCTCCACGCTTCATCCA 1419
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
|||||
1420 CGCGTGTCCCTCCT.....CCTCCAGCAGTGTCCAGGGGGCTGCT 1460
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
|||||
1461 CCCTATAGTCTCTTCATGCGAGCTCCGAGCAGGAGATGGTACAAGTGT 1510
411 eileProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisI 428
|||||
1511 CATCCCGCCCGAGCTGTGGCGCCATCATTTGGCAAGAAAGGGCCAGCA 1560
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
|||||
1561 TCAAAACACTCTCCCGCTTCGCCAGCGCTCCATCAAGATTCACACCACA 1610
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProGl 461
|||||
1611 GAAACACTGACTCCAAAGTTCGAATGCTGCTCATCTACTGGACCCCCAGA 1660
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluGluA 478
|||||
1661 GGCTCAGTTCAAGGCCAGGAGAAATCTATGGCAAACTAAAGAAAGAGA 1710
478 snPheValSerProLysGluValLysLeuGluAlaHisIleArgVal 494
|||||
1711 ATTTCTTTGGTCCCAAGGAGGAAGTAAGCTAGAGACCCACATACGGT 1760

```


495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAs 511
||||: |||||||||
1761 CCGGCTTCAGACGCGCGTGTCTCGCAAGGCGGCAAAAGGTGAA 1810
511 nGluLeuGlnAsnLeuSerSerAlaGluValValProArgAspGlnT 528
||||: |||||||||
1811 TGAGCTGCAGAACTTTCAGCGCAGCTAGTGTAGTCCCAAGAGACCAGA 1860
528 hrProAspGluAsnAspGlnValValLysIleThrGlyHisPheTyr 544
||||: |||||||||
1861 CCGCGATCAGAACGACCAAGTCATTTGTAAGATCATCGACATTTCTAT 1910
545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLy 561
||| |||: |||||||||
1911 GCCAGCCAGATGGCTCAGCGGAAGATCCGACAGATCCTGGCTCAAGTTAA 1960
561 sGlnHisGlnGlnGlnLysAlaLeuGlnSerGlyProGlnSerArgA 578
||| |||: |||||||||
1961 GCAA...CAGCACCAGAGGGA...CAGAGCAACCTGGCCCGAGGCTCGGA 2004
578 rGLys 579
|||||
2005 GGAAG 2009

seq_name: gb_est2:BF608004

seq_documentation_block: 769 bp mRNA EST 01-APR-2001
LOCUS BF608004
DEFINITION MY1_000946 Mouse 9-day fetus cDNA library ICRP522 Mus musculus
CDNA clone ICRFP522M2029 5', mRNA sequence.

ACCESSION BF608004
VERSION BF608004
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 769)
AUTHORS Yahyawi, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
Lehrach, H. and O'Brien, J.
TITLE Detection of a high number of novel genes in a 9-day mouse embryo
cDNA library normalised by oligonucleotide fingerprinting

JOURNAL Unpublished (2001)
COMMENT Contact: Hennig S
Laboraty 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de

EST's are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-GAGCTATTCCAGAAGTAGTGA-3'
BACKWARD: 5'-TAATACGACTCACTATAGG-3'
Seq primer: 5'-ATTAGGTGACACTATAG-3'
High quality sequence stop: 769.

FEATURES
Location/Qualifiers
1..769
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRFP522M2029"
/clone_lib="Mouse 9-day fetus cDNA library ICRP522"
/tissue_type="whole embryo"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/note="Vector: pSVSPORT1; Site 1: NotI; Site 2: SalI;
Library preparation by oligo_dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."

BASE COUNT 208 a 195 c 191 g 174 t 1 others
ORIGIN

alignment_scores:
Quality: 1155.00 Length: 242
Ratio: 4.915 Gaps: 1
Percent Similarity: 97.107 Percent Identity: 94.628
alignment_block:
US-09-685-696-176 x BF608004 ..
Align seg 1/1 to: BF608004 from: 1 to: 769
316 GlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAs 332
||||: |||||||||
45 CAGGAACCTGACGCTGTACATATCGGAACGACCATACAGTGAAGGCAG 94
332 nValGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgG 349
: |||||||||
95 TGTGTGAGACGTGTGCCAAGCGGAGGAGGAATAATAAGAAAGATCAGG 144
95 TGTGTGAGACGTGTGCCAAGCGGAGGAGGAATAATAAGAAAGATCAGG 144
349 LuSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIle 365
||||: |||||||||
145 AGTCTTATGAAATGATATTGCTTCCATGAATCTTCAAGCACATTAAATC 194
366 ProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMe 382
||||: |||||||||
195 CCTGGATTAAATCTGAATGCTTTGGGTCTTCCACCCACGTCAGGGAT 244
382 tProProProThrSerGlyProProSerAlaMetThrProProTyrProG 399
||||: |||||||||
245 GCACACTCCACCTCAGGCGCCCTTCAACCTGACTCTCCCTACCCAC 294
399 InPheGluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeu 415
||||: |||||||||
295 AATTGAGCAATCAGACGAGACTGTGCATCTGTTATTCGCGCCTG 344
416 SerValGlyAlaIleLeGlyLysGlnGlyGlnHisIleLysGlnLeuSe 432
||||: |||||||||
345 TCCGTTGGGCGCCATCATTTGCCAAGCAGGCGGCAACACATCAACAGCTTC 394
432 rArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspA 449
||||: |||||||||
395 TCCTTTGCGGGAGCTTCGATTAAAGATCGCTCCAGCGAAGCACCAGATG 444
449 lAlysValArgMetValIleIleThrGlyProProGluAlaGlnPheLys 465
||||: |||||||||
445 CTAAAGTCCGATGGTGGATTTATCCTGGACACACAGAGGCTCAGTTCAAG 494
466 AlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerPr 482
||||: |||||||||
495 GCTCAGGGAAGAATTTATGCAAAATTTAAAGAAACACACTTTGTTAGTCC 544
482 oLysGluGluValLysLeuGluAlaHisIleArgValProSerPheAlaA 499
||||: |||||||||
545 TAAAGAAAGAGGTGAACTTTGAAGCTCAGATGCGCGCTCTTCTGCTG 594
499 lAGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsn 515
||||: |||||||||
595 CTGCGCAGATTATTGGGAAAGAGGCAAAACTGTGAATGAGCTCCAGAGT 644
516 LeuSerSerAlaGluValValProArgAspGlnThrProAspGluAs 532
||||: |||||||||
645 TTATCAGCTGTTGAGTGTGCTCCCGCGTACCAGACACCTGATGAGAA 694
532 nAspGlnValValLysIleThrGlyHisPheTyrAla.CysGlnVal 548
||||: |||||||||
695 TGATCAAGTAGTTGTCAACATAAATGCGCCACTTCTATGCTTTCGCCGGT 744
549 AlaGlnArgLysIleGlnGluIle 556
||||: |||||||||
745 GCCCAGAGGAAATTCAGGAAATT 768
seq_name: gb_est1:AU127673

```

seq_documentation_block:
  LOCUS       AU127673       733 bp      mRNA           EST           24-OCT-2000
  DEFINITION   AU127673 NT2RP2 Homo sapiens cDNA clone NT2RP2001818 5', mRNA
  sequence.
  ACCESSION   AU127673
  VERSION     AU127673.1 GI:10988027
  KEYWORDS    EST.
  SOURCE      human.
  ORGANISM    Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1 (bases 1 to 733)
  AUTHORS     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
               Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
               Isogai,T.
  TITLE       HRI human cDNA project
  JOURNAL     Unpublished (2000)
  COMMENT     Contact: Takao Isogai
               Genomics Laboratory
               Helix Research Institute
               1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
               Tel: 81-438-52-3951
               Fax: 81-438-52-3952
               Email: genomics@hri.co.jp
               HRI human cDNA project; 5' - & 3'-end one pass sequencing: Helix
               Research Institute; cDNA library construction: Department of
               Virology, Institute of Medical Science, University of Tokyo, and
               Helix Research Institute.
  FEATURES    Location/Qualifiers
               1..733
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="NT2RP2001818"
               /clone_lib="NT2RP2"
               /cell_type="teratocarcinoma"
               /cell_line="NT2"
               /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
               cells after 2-weeks retinoic acid (RA) induction"
  BASE COUNT  212 a 181 c 148 g 180 t 12 others
  ORIGIN

alignment_scores:
  Quality: 1065.50      Length: 230
  Ratio: 4.910          Gaps: 2
  Percent Similarity: 94.348 Percent Identity: 93.913

alignment_block:
  US-09-685-176 x AU127673 ..
  Align seg 1/1 to: AU127673 from: 1 to: 733

194 CysAspLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaI 210
|||||
5 TGTGATTGGCTCTCGCCCTGCTGGTTCACCAACCAATTTGTTGGAGCAT 54
|||||

210 eileGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnS 227
|||||
55 CATAGGAAGAAGAGTCCACCATTCGGACATCCCAACACAGACCCAGT 104
|||||

227 erTysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSer 243
|||||
105 CTAATAATCGATGTCACCGTAACAAAATGCGGGGCTGCTGAGAACTCG 154
|||||

244 IleThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerI 260
|||||
155 ATTAATACTCTCTACTCTCTGAAGGACCTCTCGCGCTTTGTAAGTCTAT 204
|||||

260 eLeuGluIleMethIstLysGluAlaGlnAspIleLysPheThrGluGlu 277
|||||
205 TCTGGAGATTATGCTAAGGAGCTCAAGATATAAAAATTCACANAAGAGA 254
|||||

```

```

277 leProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGly 293
|||||
255 TCCCTTTGAAGATTTAGCTCATAATAACTTTGTTGGACGCTTATGGT 304
|||||

294 LysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysI 310
|||||
305 AAAGAAGGAAGAAATCTTAAAAAATTGAGCAAGACACAGACNCTAAAT 354
|||||

310 eThrIleSerProLeuGlnGluLeuThrLeuTyAsnProGluArgThrI 327
|||||
355 CACGATATCTCCATTGCAGGAATTGACGCTGTATTAATCCAGAACGCACTA 404
|||||

327 leThrValLysGlyAsnValGluThrCysAlaLysAlaGluGluIle 343
|||||
405 TTACAGTTAAAGGCAATGTTGAGACATCTGCCAAAGCTGAGGAGGATC 454
|||||

344 MetLysLysIleArgGluSerTyGluAsnAspIleAlaSerMetAsnI 360
|||||
455 ATGAAGAANATCAGGGAGTCTTATGAAAATGATATTGCTTCTATGAATCT 504
|||||

360 uGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe 377
|||||
505 TCNAGCACATNTAATTCCTGNATTAAATCTGAAGCCCTGGGTCTGTTC 554
|||||

377 roProThrSerGlyMetProProThrProThrSerGlyProProSerAlaMet 393
|||||
555 CACCACCTTCAGGGATGCACCTCCACCTCCAGGCCCTTCAGCCATG 604
|||||

394 ThrProProTyProGlnPheGluGlnSerGluThrGluThr.ValHISG 410
|||||
605 ACTCCTCCCTCCGCCAGTTTGAGCAATCCGAAACGGAGACTGGTTCNTC 654
|||||

410 lnPheIleProAlaLeu...SerValGlyAlaIleIle 421
|||||
655 TGTTTTATCCCCACCTCTATCCAGTCGTCGCATCATC 692

seq_name: gb_est1:BE545535

seq_documentation_block:
  LOCUS       BE545535       790 bp      mRNA           EST           09-AUG-2000
  DEFINITION   601070391F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456407 5',
  mRNA sequence.
  ACCESSION   BE545535
  VERSION     BE545535.1 GI:9774180
  KEYWORDS    EST.
  SOURCE      human.
  ORGANISM    Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1 (bases 1 to 790)
  AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
  COMMENT     Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-remail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               plate: LLAM8444 row: f column: 24
               High quality sequence stop: 667.
  FEATURES    Location/Qualifiers
               1..790
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3456407"
               /clone_lib="NIH_MGC_12"
               /tissue_type="cervical carcinoma cell line"
               /lab_host="DH10B"
               /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;

```

Site_2: Sali; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 221 a 193 c 209 g 167 t
ORIGIN

alignment_scores:

Quality: 1047.50 Length: 259
Ratio: 4.476 Gaps: 4
Percent Similarity: 90.347 Percent Identity: 85.714

alignment_block:

US-09-685-696-176 x BE545535 ..

Align seg 1/1 to: BE545535 from: 1 to: 790

```
29 valSerGlyProPheLeuValLysThrGlyTyrAlaPheValAspCysPr 45
|||||
3 GTGTCGGGACCTTCTGTTGAAGACTGGCTAGCGTTCGTGACTGCC 52
|||||
45 oAspGluSerTrpAlaLeuLysAlaIleGluAlaLeuSerGlyLysIleG 62
|||||
53 GGACGAGAGCTGGCCCTCAAGCCATCGAGGCGCTTCAGGTAAATAG 102
|||||
62 luLeuHisGlyLysProIleGluValGluHisSerValProLysArgGln 78
|||||
103 AACTGCACGGAAACCATAGAGTTGAGCACTCGTCCCAAAAGGCAA 152
|||||
79 ArgIleArgLysLeuGlnIleArgAsnIleProProHisLeuGlnTrpGl 95
|||||
153 AGGATTCGGAACCTTCAGATACAAATATCCCGCCTCATTTACAGTGGGA 202
|||||
95 uValLeuAspSerLeuValGlnTyrGlyValValGluSerCysGluG 112
|||||
203 GGTCTGGTAGTTTACTAGTCAGATGCGAGTGGTGGAGAGCTGTGAGC 252
|||||
112 lnValAsnThrAspSerGluThrAlaValValAsnValThrTyrSerSer 128
|||||
253 AACTGACACTGACTCGGAACCTGCAGTTGTAATGTAACTATTTCAGT 302
|||||
129 LysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGl 145
|||||
303 AAGGACCAAGCTAGACAAGCACTAGACAACACTGAATGGATTTTCAGTTAGA 352
|||||
145 uAsnPheThrLeuLysValAlaTyrIleProAspGluMetAlaAlaGlnG 162
|||||
353 GAATTTACCTTGAAAGTAGCCCTATATCCCTGATGAATGGACGCCAGC 402
|||||
162 lnAsnProLeuGlnGlnProArgGlyArgGlyLeuGlyGlnArgGly 178
|||||
403 AAAACCCCTTCGACGAGCCCGGAGTGCCTGGGGCTTGGGACAGGGGC 452
|||||
179 SerSerArgGlnGlySerProGlySerValSerLysGlnLysProCysAs 195
|||||
453 TCCTCAAGCAGGGTCT.CCAGGATCGGTATCCAAAGCAGAAACCATGTGA 501
|||||
195 pLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaIleIle. 211
|||||
502 TTTGCTCTCTGCGCTGCTGTTGCCACCCCAATTTGTTGGAGCCATCATAG 551
|||||
212 GlyLysGluGlyAlaThrIleArgAsnIle.ThrLysGlnThrGlnSerL 228
|||||
552 GGAAGAAGAGTGCACCACTTCGGAACATCTACCAACAGACCCAGTCTA 601
|||||
228 ysIleAspVal.HisArgLysGluAsnAlaGlyAlaAlaGluLysSerIl 244
|||||
602 AAATCATGTGCCACCGTGAAGAAATCGGGCTGCTG...AGAAGTCG 648
|||||
244 eThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleL 261
|||
649 ACTACTATCTCTCTATCTGAAGGAGCCTCTGGGATTTGTAAGACTATCCT 698
|||
```

261 euGluIleMetHisLysGluAlaGlnAspIleLysPheThrGluGluIle 277
|||||
699 GGGAATATTGCTAAGGGAGCGCAGATTTTAAATCCAGAGAGAATC 748
|||||

278 ProLeuLysIleLeuAlaHis 284
|||||
749 CCTGTAAATTTTATGCCCAT 769
|||||

seq_name: gb_est1:AL584627

seq_documentation_block:

LOCUS AL584627 707 bp mRNA EST 28-FEB-2001
DEFINITION AL584627 Stratagene Chick Embryo Lambda cDNA Library (* 937405)
Gallus gallus cDNA clone ROS012C05, mRNA sequence.

ACCESSION AL584627
VERSION AL584627.1 GI:13163358
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 707)
Murray,F.
Stratagene Chick Embryo Lambda cDNA Library

AUTHORS

TITLE

JOURNAL

COMMENT

Dept. Genomics and Bioinformatics

Roslin Institute

Roslin, Midlothian, EH25 9PS, UK

Tel: +44 (0)131 527 4200

Fax: +44 (0)131 440 0434

Email: frazer.murray@bbsrc.ac.uk

Seq primer: F3.

Location/Qualifiers

1..707

/organism="Gallus gallus"

/db_xref="taxon:9031"

/clone="ROS012C05"

/clone_lib="Stratagene Chick Embryo Lambda cDNA Library (*

937405)"

/tissue_type="Embryo"

/dev_stage="5 days old"

/lab_host="SOLR cells (kanamycin resistant)"

/note="Vector: pBLUESCRIPT SK; Site_1: EcoRI; Site_2: XhoI

; Cloned unidirectionally. Primer: Oligo dt. Uni-ZAP XR

vector. Average insert size: 1.5kb.; 5' adaptor sequence:

5' GAATTCGCGCAGAG 3'; 3' adaptor sequence: 5'

CTCGAGTTTCTTTTCTTTT 3'

BASE COUNT 218 a 150 c 159 g 180 t

ORIGIN

alignment_scores:

Quality: 975.50 Length: 231
Ratio: 4.495 Gaps: 4
Percent Similarity: 93.939 Percent Identity: 85.714

alignment_block:

US-09-685-696-176 x AL584627 ..

Align seg 1/1 to: AL584627 from: 1 to: 707

298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314

|||||

17 AACCTGAAGAAATTTGACGAGGACACAGATACTAAATACGATATCTCC 66

|||||

314 oLeuGlnGluThrLeuTyrAsnProGluArgThrIleThrValLysG 331

|||||

67 ATTGCAGGACTTGACACTCTATATCCGGAAGGACCATACAGTTAAG 116

|||||

331 lyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLysIle 347

|||||

```
117 CAGTATTGAACCTTGTGCCAAGCCGAGGAAGAAATTTATGAAGAAATC 166
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
167 AGGGAATCCTATGAAATGATATTGCTGCTATGAATCTTCAAGCACATT 216
364 uileProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProPro...Thrs 380
217 AATTCCTGGATTAAATCTGAATGCTTGGTCTGTTCCACACTCTCTCT 266
380 erGlyMetProProPro...ThrsGlyProProSerAlaMetThrPro 395
267 CAGGAATACCACTTCTCCAGCTGAGTGTGCTCTGCTGCTGCTGCT 316
396 ProTyrProGlnPhe...GluGlnSerGluThrGluThrValHisGlnPh 411
317 TCATATCGCCATTGGGCAACACGCCGAGTCTGAGACTGTTTCATCTGT 366
411 eileProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyHisI 428
367 CATCCAGCCTTGGCAGTTGAGCTATTATTGGCAAGCAGGACAGACA 416
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
417 TCAACAACACTTTCGCTTTGCAGCGGATCTATTAGATTGCCCTGCT 466
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProG 461
467 GAAGCAGCAGATGCCAAGCTGAGAATGTTATCATCACTGCACTCCAGA 516
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluAla 478
517 AGCTCAGTTTAAAGCTCAAGGGAGAACTTATGGTAAAGGAGGCAAAAG 566
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
567 ATTCTTTGGACCTTAAGAAAGAGTGAACCTTGAGGCTCATATAAAGTG 616
495 ProSerPheAlaLagArgValIleGlyLysGlyGlyLysThrValas 511
617 CCATCTTATGCTGCTGGTAGAGTTATTGGTAAAGGAGGCAAAAGTAA 666
511 nGluLeuGlnAsnLeuSerSerAlaGluValValPro 524
667 TGAGCTTCAAAACTTGACAAGTGGCAGAAGTTGTAGTCCCT 707
```

seq_name: gb_est2:BG079106

```
seq_documentation_block: 587 bp mRNA EST 26-JAN-2001
LOCUS BG079106
DEFINITION H3035G11-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG079106
VERSION BG079106
KEYWORDS BG079106.1 GI:12561674
SOURCE EST.
ORGANISM house mouse.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 587)
AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
T.S., Carter,M.G. and Ko,M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTS: H3035G11-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3035 row: G column: 11
```

Seq primer: -21M13 Reverse
High quality sequence stop: 587
POLYA=No.

FEATURES

Location/Qualifiers
1..587
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3035G11"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"

/lab_host="DH10B"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 156 a 152 c 135 g 143 t 1 others
ORIGIN

alignment_scores:
Quality: 952.00 Length: 195
Ratio: 5.011 Gaps: 0
Percent Similarity: 97.436 Percent Identity: 94.872
alignment_block:
US-09-685-696-176 x BG079106 ..
Align seg 1/1 to: BG079106 from: 1 to: 587
353 AsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeuAs 369
1 AATGATATTGCTTCCATGAATCTTCAAGCACATTTAATCCCTGGATTAA 50
369 nLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProProProt 386
51 TCTGAATGCCCTTGGTCTGTTCACCCACCCAGGATGACCCGCCCA 100
386 hrSerGlyProProSerAlaMetThrProProTyrProGlnPheGluGln 402
101 CCTCAGGCCCCCTTCAACCTGACTCTCTCCCTACCCACAATTTGAGCAA 150
403 SerGluThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAl 419
151 TCAGAGACGGAGACTGTGCATCTGTTATTCCCGCCCTGTCCGTGGCGC 200
419 aIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaG 436
201 CATCTTGGCAAGCGGCCCAACACATCAACACAGCTTCTCGCTTGGCG 250
436 lyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysValArg 452
251 GAGCTTCGATTAAAGATCGCTCCAGCCGGAAGCAGACGATGCTAAAGTGG 300
453 MetValIleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyAr 469
|||||

301 ATGGTGATTATCACTGGACCACAGAGGCTCAGTTCAAGGCTCAGGAAG 350
469 gletyrglylyllelysglucluasnphervalserprolysgluclu 486
|||||
351 AATTTATGAAAAATTAAGAGAAATTTGTTAGTCTCAAGAGAGG 400
486 allysleucluhalahisileargvalproserphealaalaglyargval 502
|||||
401 TGAACCTTGAAGCTACATACAGAGTCCCTCTTCTGCTGGCATAGTT 450
503 ileglylysglyglythyrvalasnglucluasnleuasnleuseral 519
|||||
451 ATTGCAAGAGGAGGCAAAACGCTGATGAGCTCCGGAGTTTATCAAGTCC 500
519 agluvalvalproargaspclnthrproaspgluasnaspclnvalv 536
|||||
501 TGAAGTTGTCGACCCCGTCACAGACACCTGATGAGATGATCAAGTTG 550
536 alvalysllethrghlyhisphetyrAlaCysGln 547
|||||
551 TTGTACACATACTGGCCACTTCTATGCTTGCCAG 585

seq_name: gb_estl:AA117282

seq_documentation_block: 677 bp mRNA EST 15-NOV-1996
LOCUS AA117282
DEFINITION mn25a10.r1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:538938 5' similar to gb:L35599 Mus musculus Y-box
binding protein mRNA, 3' end (MOUSE);, mRNA sequence.
ACCESSION AA117282
VERSION AA117282.1 GI:1672303
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 677)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:325874

Seq primer: -40ml3 ET

High quality sequence stop: 411.

Location/Qualifiers

source

1. 677
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:538938"
/clone_lib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/note="Organ: whole embryo; Vector: pCMV-SPORT; Site:1;
Sali; Site:2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Gastrulating embryos were collected at 7.5dpc
from C57BL6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).

BASE COUNT 196 a 171 c 152 g 158 t
ORIGIN
alignment_scores:
Quality: 936.50 Length: 219
Ratio: 4.591 Gaps: 2
Percent Similarity: 93.151 Percent Identity: 86.758

alignment_block:

US-09-685-696-176 x AA117282

Align seg 1/1 to: AA117282 from: 1 to: 677

215 GlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspVa 231
|||||
1 GGTGCTACCATTCGCAACATCACAAAGCAGACCCAGCTCTAAATCGATGT 50
231 lHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuS 248
|||||
51 CCATCGTAAGGAGATACAGGGCGCGGAGAGATGCATTACTATCTCT 100
248 erThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMet 264
|||||
101 CTACCCCTGAAGGCACCTCTGCAGCTTGTAAGTCTATTCTGGAGATTATG 150
265 HisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIl 281
|||||
151 CATAGAAGAGCTCAAGATATAAAATTCACAGAGAGATTCCTCTGAAGAT 200
281 eLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgA 298
|||||
201 CTTAGCTCAATAAATTTGTAGCCGCTCTCATTGGTAAAGAAGAGAA 250
298 snLeuLysLysIleGluGlnAspThrAspThrLysLysIleThrIleSerPro 314
251 ACCTTAAAAAATTCGACGACAGACGACACTAAATATCAACAATATCTCCA 300
315 LeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGl 331
|||||
301 TTGCAGAACTGACGCTGTACAAATCCGGAACGCCACCATTTACAGTGAAGG 350
331 yAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleA 348
|||||
351 CACTGTTGACAGCTGTGCCAAGCGGAGGAGAGAAATAATGAAGAAGATCA 400
348 rgGluSerTyrGluAsnAspIleAlaSerMetAsn.LeuGlnAlaHisLe 364
|||||
401 GGGAGTCTTATGAAATGATATTGCTTCCATGAATCCTTCAAGCACATTT 450
364 uileProGlyLeuAsnLeuAlaLeuGlyLeuPheProThrSerG 381
|||||
451 AATCCCTGGATTAAATCTGAATGGCTGGTGGTCTGGTCCCACCGTCAGG 500
381 lyMetProProThrSerSerGlyProProSerAlaMetThrProProTyr 397
|||||
501 GATGCCACTTCCACCTCAGGGGCCCTTCAACCCCTGACTCTCTCCCTAC 550
398 ProGlnPheGluGlnSerGluThrGluThrValHisGlnPheIleProAl 414
|||||
551 CCACAATTTGGCAATCAGACGAGGGGACTGTGCATCTGTTTATTTCCTCCG 600
414 aLeuSerValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnL 431
|||||
601 CCTGTGTTGGCGCAATCATTTGGCA...GCCGGGCCACACATTAACAGCT 647
431 euSer 432
:||||
648 TTTCG 652

seq_name: gb_estl:AA163596

seq_documentation_block: 645 bp mRNA EST 17-DEC-1996
LOCUS AA163596
DEFINITION mm38g02.r1 Beddingington mouse embryonic region Mus musculus cDNA
clone IMAGE:540242 5' similar to gb:L35599 Mus musculus Y-box
binding protein mRNA, 3' end (MOUSE);, mRNA sequence.
ACCESSION AA163596
VERSION AA163596.1 GI:1739542
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 645)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:327178
Seq primer: -40m13 ET
High quality sequence stop: 465.
Location/Qualifiers
1. .645
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:540242"
/clone_lib="Beddingington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
Sali; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Gastrulating embryos were collected at 7.5dpc
from C57BL6 x DBA matings, excluding embryos that had
developed head folds and all extraembryonic tissues.
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."
BASE COUNT 178 a 159 c 150 g 157 t 1 others
ORIGIN

alignment_scores:
Quality: 911.00 Length: 218
Ratio: 4.401 Gaps: 3
Percent Similarity: 94.954 Percent Identity: 92.202

alignment_block:
US-09-685-696-176 x AA163596 ..

Align seg 1/1 to: AA163596 from: 1 to: 645

349 GluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeu11 365
|||||
3 GAGTCTTATGAAATGATATGCTTCATGAAATCTTCAGCACATTTAAT 52

365 eProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyM 382
|||||
53 CCCTGGATTAATCTGAATGCTTGGGTCTCTCCACCCAGCGCAGGA 102

382 etProProProThrSerGlyProProSerAlaMetThrProProTyrPro 398

|||||
103 TGGCACCTCCACCTTCAGGGCCCCCTTCAACCCCTGACTCCTCCCTACCCA 152
|||||
399 GlnPheGluGlnSerGluThrGluThrValHisGlnPheIleProAlaLe 415
|||||
153 CAATTTGAGCAATCAGACGAGACTGTCATCTGTTTATTCCGCCCT 202
|||||
415 uSerValGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuS 432
|||||
203 GTCCGTTGGCGCCATCATTTGCAAGCAGGGCCACACATCAACAGCTTT 252
|||||
432 eArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAsp 448
|||||
253 CTCGCTTTGC. GGAGCTTCGATTAAAGATTCGCTCCAGCCGAAGACCAGAT 301
|||||
449 AlaLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLy 465
|||||
302 CTTAAAGTCGGATGGGTGATTATCACTGGACACACAGAGGCTCAGTTCAA 351
|||||
465 sAlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerP 482
|||||
352 GGCTCAGGGAAGATTTATGAAAAAATTAAGAGAGAAACTTTGTAGTC 401
|||||
482 rOLysGluGluValLysLeuGluAlaHisIleArgValProSerPheAla 498
|||||
402 CTTAAAGAGAGGTGAACCTTGAAGCTCACATCAGAGTCCCGCTCTTGTCT 451
|||||
499 AlaGlyArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAs 515
|||||
452 GCTGCAGAGTATTGGAAAGAGGCAAAACGGTGAATGAGCTCCAGAG 501
|||||
515 nLeuSerSerAlaGluValValProArgAspGlnThrProAspGluA 532
|||||
502 TTTATCAAGTGCTGAAGTTGCGT. CCCCGTGACCAGACACCTGATGAGA 550
|||||
532 sNAspGlnValValLys. lIeThrGlyHisPheTyrAlaCysGlnVa 548
|||||
551 ATGATCAAGTAGTTGTCAATAAAC...TGGCACTTCTATGCTGCTA.GT 596
|||||
548 lAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnG 565
|||||
597 TGCCACAGAGANATTTCAGA...ATTCGACTCAGTAAGCAGCAGCAGCAGC 643
|||||
565 ln 565
||
644 AG 645

seq_name: gb_estl:AL584575

seq_documentation_block: 647 bp mRNA EST 28-FEB-2001
LOCUS AL584575
DEFINITION AL584575 Stratagene Chick Embryo Lambda cDNA Library (* 937405)
Gallus gallus cDNA clone ROS011B11, mRNA sequence.
ACCESSION AL584575
VERSION AL584575.1 GI:13163306
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 647)
AUTHORS Murray, F.
TITLE Stratagene Chick Embryo Lambda cDNA Library
JOURNAL Unpublished (2001)
COMMENT Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
Seq primer: r3.

FEATURES
source

```
Location/Qualifiers
1. .647
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="ROS011B11"
/clone_lib="Stratagene Chick Embryo Lambda cDNA Library (*)
937405)"
/tissue_type="Embryo"
/dev_stage="5 days old"
/lab_host="SOLR cells (kanamycin resistant)"
/notes="vector: pBLUESCRIPT SK; Site_1: EcoRI; Site_2: XhoI
vector. Cloned unidirectionally. Primer: Oligo dt. Uni-ZAP XR
vector. Average insert size: 1.5kb.; 5' adaptor sequence:
5' GAATTCGGCAGCAG 3'; 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'."
```

BASE COUNT 202 a 139 c 142 g 164 t

ORIGIN

alignment_scores:

Quality: 909.50 Length: 211
Ratio: 4.593 Gaps: 3
Percent Similarity: 93.839 Percent Identity: 86.256

alignment_block:

US-09-685-696-176 x AL584575 ..

Align seg 1/1 to: AL584575 from: 1 to: 647

```
280 LysileLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGlu1 296
15 AAGATATTGGCAGACACAACTTTGTTGTCGACTTATTGGCAAGG 64
296 YATGAsnLeuLysLysIleGluGlnAspThrLysIleThrIles 313
65 AGAAGACCTTGAGAAATTTGAGCAGACACAGATCTAAATCCACGATAT 114
313 erProLeuGlnGluLeuThrLeuTyAsnProGluArgThrIleThrVal 329
115 CTCCATTGAGGACTTGACACTCTATAATCCGGAAGGACCATACAGTT 164
330 LysGlyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLy 346
165 AAAGGCGATTGAAACTTTGTGCCAAGCGCAGAGAAATTTATGAAGAA 214
346 sileArgGluSerTyArgLeuAsnAspIleAlaSerMetAsnLeuGlnAlaH 363
215 AATCAGGGAATCTATGAAATGATATGCTGCTATGAATCTTCAAGCAC 264
363 isLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPhePro... 378
265 ATTTAATCTCGATTAAATCTGAATGCCTTGGTCTGTGCCACCTTCT 314
379 ThrSerGlyMetProPro...ThrSerGlyProProSerAlaMetTh 394
315 TCTTCAGGAATACCACTCTCTCGAGTGAGTGTGCTCTGCTGCTGCTG 364
394 rProProTyProGlnPhe...GluGlnSerGluThrGluThrValHisG 410
365 TGCTTCATATCCACCAATTTGGGCAACAGCCGAGTCTGAGACTGTTCA 414
410 lnPheIleProAlaLeuSerValIleIleIleGlyLysGlnGlyGln 426
415 TGTTTCATCCAGCCTTGGCAGTTGGAGCTATTATTGGCAAGCAGGACAG 464
427 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaPr 443
465 CACATCAAAACAACCTTCTCGCTTGCAGGCGCATCTATTAAAGATTGCC 514
443 oAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyProp 460
515 TGCTGAAGGACCAAGATGCCAAGCTGAGAATGGTTATCATCACTGGACCTC 564
```

```
460 roGluAlaGlnPheLysAlaGlnGlyArgIleTyrglyLysIleLysGlu 476
565 CAGAAGCTCAGTTTAAGGCTCAAGGGAGGAATCTATGGAAGAACTTAAAGAA 614
477 GluAsnPheValSerProLysGluGluValLys 487
615 GAAATTTCTTGGACCTAAAGAAGAAGTGAAA 647
```

seq_name: gb_est1:AV612377

seq_documentation_block:

LOCUS AV612377 586 bp mRNA EST 30-AUG-2000
DEFINITION AV612377 Bos taurus lung fetus Bos taurus cDNA clone EIL0045B11 5',
mRNA sequence.

ACCESSION AV612377

VERSION AV612377.1 GI:9748047

KEYWORDS EST.

SOURCE

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 586)

AUTHORS Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitohzono,A. and
Suzuki,H.

TITLE bovine cDNA sequencing

JOURNAL Unpublished (2000)

CONTACT: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@coo.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source

1. .586

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone="EIL0045B11"

/clone_lib="Bos taurus lung fetus"

/tissue_type="lung"

/dev_stage="fetus"

/lab_host="DH10B"

/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A

was deleted from a NotI site"

BASE COUNT 161 a 146 c 150 g 128 t 1 others

ORIGIN

alignment_scores:

Quality: 903.00 Length: 193
Ratio: 4.778 Gaps: 0
Percent Similarity: 97.927 Percent Identity: 92.746

alignment_block:

US-09-685-696-176 x AV612377 ..

Align seg 1/1 to: AV612377 from: 1 to: 586

```
114 AsnThrAspSerGluThrAlaValAlaValThrTyrSerSerLysAs 130
:
:
:
2 CAGCGTCCCGGAAACTGCTGTGTCACGTAACCTATTCCAATAGGA 51
```

```
130 pGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsn 147
:
:
:
52 GCAAGCCAGACAAGCTTTAGACAAGCTGAATGGTTCAGCTGGAGAACT 101
```

```
147 heThrLeuLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsn 163
:
:
:
102 TCACACTGAGAGTGGCCTACATCCCCGATGAGATGGCGCCAGCAGAGC 151
```


546 CysGlnValAlaGlnArgLysIleGlnGluIle.LeuThrGlnValLysG 562
 :: :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 693 AGCTCAGATGGCTCACGGAAGATCCGAGAACTTCCTGGCCCGGTTAAGC 742
 562 inHisGlnGlnLysAlaLeuGlnSerGlyProGln 575
 743 CG...CGACGATCGAAAGACGGGTTCCTCCCGCCCGCGGAG 780

seq_name: gb_est1:AUI34323

seq_documentation_block: 745 bp mRNA 24-OCT-2000
 LOCUS AUI34323 OVARC1 Homo sapiens cDNA clone OVARC1001723 5', mRNA
 DEFINITION AUI34323 OVARC1 Homo sapiens cDNA clone OVARC1001723 5', mRNA
 sequence.
 ACCESSION AUI34323
 VERSION AUI34323.1 GI:10994862
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 745)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.

TITLE HRI human cDNA project

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES Location/Qualifiers

source 1..745
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="OVARC1001723"
 /clone_lib="OVARC1"
 /tissue_type="ovary, tumor tissue"
 /note="Vector: pME18SFL3"
 BASE COUNT 232 a 168 c 170 g 172 t 3 others
 ORIGIN

alignment_scores:
 Quality: 884.00 Length: 177
 Ratio: 5.023 Gaps: 0
 Percent Similarity: 99.435 Percent Identity: 99.435

alignment_block:

US-09-685-696-176 x AUI34323 ..

Align seg 1/1 to: AUI34323 from: 1 to: 745

403 SerGluThrGluThrValHisGlnPheIleProAlaLeuSerValGlyAl 419
 |||||
 3 TCAGAACGAGACTGTTCTATCTGTTTCCAGCTATCATCGTCGTGC 52

419 alIleIleGlyLysGlnGlyHisLeuLysGlnLeuSerArgPheAlaG 436
 |||||

53 CATCATCGGACGAGCGGCGAGCACATCAAGCAGCTTTCGCTTTGCTG 102

436 lylalaSerlleLyslleAlaProAlaGluAlaProAspAlaLysValArg 452
 |||||

103 GAGCTTCAATTAAGATTGTTCCAGCGGAGCACCATGCTAAAGTGAGG 152

453 MetValIleIleThrGlyProGluAlaGlnPheLysAlaGlnGlyAr 469
 |||||
 153 ATGGTGATTATCACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAG 202
 469 gileTyrglyLysIleIysGluGluAsnPheValSerProLysGluGluV 486
 |||||
 203 AATTATGGAAAAATTAAGAAGAAACTTTGTAGTCTTAAAGAAGAGG 252
 486 alLysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgVal 502
 |||||
 253 TGAACCTTGAAGCTCATATCAGAGTGCCATCCTTTGCTGCTGCAGAGTT 302
 503 IleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSerAl 519
 |||||
 303 ATTGGAAGAGGAGGCAAAACGGTGAATGAACCTTCAGAAATTTGTCAAGTGC 352
 519 aGluValValValProArgAspGlnThrProAspGluAsnAspGlnValV 536
 |||||
 353 AGAAGTTGTTGCTCCTCGTGACCAGACACCTGATGAGAATGACCAAGTGG 402
 536 alValLysIleThrGlyHisPheTyraLaCysGlnValAlaGlnArgLys 552
 |||||
 403 TTGTCAAAATAACTGGTCACCTTCTATGCTTGCCAGGTGGCCAGAGAAA 452
 553 IleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLysAlaLe 569
 |||||
 453 ATTGAGGAATTTCTGACTCAGGTAAAGCAGCACCAACAACAGAGGCTCT 502
 569 uGlnSerGlyProGlnSerArgArgLys 579
 |||||
 503 GCAAAGTGGACCACTCAGTCAAGACGGAAG 533

seq_name: gb_est2:BG082113

seq_documentation_block:

LOCUS BG082113 573 bp mRNA 26-JAN-2001
 DEFINITION H3072F04-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3072F04 5', mRNA sequence.

ACCESSION BG082113

VERSION BG082113.1 GI:12564681

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 573)

AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka

,T.S., Carter,M.G. and Ko,M.S.H.

Verification and initial annotation of NIA mouse 15K cDNA clone set

Unpublished (2001)

Other ESTs: H3072F04-3

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@ngsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.

Plate: H3072 row: F column: 04

Seq primer: -21M13 Reverse

High quality sequence stop: 573

POLYA-No.

FEATURES

source Location/Qualifiers

1..573

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="H3072F04"

/clone_lib="NIA Mouse 15K cDNA Clone Set"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev_stage="Clones arrayed from a variety of cDNA libraries"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Sali; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1976."

BASE COUNT 167 a 132 c 150 g 124 t
ORIGIN

alignment_scores:
Quality: 868.00 Length: 178
Ratio: 4.932 Gaps: 0
Percent Similarity: 98.876 Percent Identity: 97.191

alignment_block:

US-09-685-696-176 x BG082113

Align seg 1/1 to: BG082113 from: 1 to: 573

402 GlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerValG1 418
2 CAATCAGACGAGGACGTGTCATCTGTTTATCCCGCCCTGTCGGTGG 51
418 yAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheA 435
52 CGCCATCATTTGGCAGGAGGCGCAACACATCAACAGCTTTCTCGCTTG 101
435 laGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLysVal 451
102 CGGAGCTTCGATTAAAGTCCCTCCAGCTCTTTCCACGAGTGTAAAGTG 151
452 ArgMetValIleIleThrGlyProGluAlaGlnPheLysAlaGlnG1 468
152 CGGATGGTGATTATCGCTGGACACCAGAGGCTCAGTTCAAGGCTCAGG 201
468 yArgIleThrGlyLysIleLysGluGluAsnPheValSerProLysGluG 485
202 AAGAAATTTATGGAATAATTAAGAGAAACCTTTGTAGTCTCTAAAGAAG 251
485 luValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArg 501
252 AGGTGAACCTTGAAGCTCACATCAGAGTCCGCTCTTTGCTGGCAGA 301
502 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuSerSe 518
302 GTTATTGGAAAGGAGGCAAAACCGTGAATGAGCTCCAGAGTTTATCAAG 351
518 rAlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnV 535
352 TGCTGAAGTTGCTGCTCCCGCTGACACACACCTGATGAGAATGATCAAG 401
535 alValValLysIleThrGlyHisPheThrAlaCysGlnValAlaGlnArg 551
402 TAGTTGTCAAAATTAACGGCCACTTCATGCTTGCACAGGTTGCCAGAGG 451
552 LysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLysAl 568
452 AAAATTCAGGAATTCGACTCAGGTAAAGCAGGACCAGCAGCAGAGAAAGC 501

568 aLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
502 TCTGCAGAGTGACACCTCAGTCAAGCGGAAG 535

seq_name: gb_estl:AA691146

seq_documentation_block:
LOCUS AA691146 559 bp mRNA EST 16-DEC-1997
DEFINITION vr68h09.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone IMAGE:1125857 5' similar to TR:000425 O00425 PUTATIVE RNA BINDING PROTEIN KOC ; mRNA sequence.

ACCESSION AA691146
VERSION AA691146.1 GI:2692082
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 559)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lannon,G., Soares,B., Wilson,R. and Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

TITLE
JOURNAL
COMMENT

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:615193
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 440.

FEATURES
source
Location/Qualifiers
1..559
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:1125857"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pBluescribe (modified); Site_1: MluI; Site_2: Sali; Cloned unidirectionally from mRNA prepared from 13,500 2-cell stage embryos. Primer: Sali(dT): 5'-CGTGCAGCGTCGACGTTTATTTT-3', cDNAs were cloned into the MluI/Sali sites of a modified pBluescribe vector using commercial linkers (NEB). Average insert size: 1.2 kb."

BASE COUNT 154 a 139 c 155 g 111 t
ORIGIN

alignment_scores:
Quality: 835.50 Length: 188
Ratio: 4.720 Gaps: 2
Percent Similarity: 94.149 Percent Identity: 89.894

alignment_block:

US-09-685-696-176 x AA691146

Align seg 1/1 to: AA691146 from: 1 to: 559

44 CysProAspGluSerTrpAla.LeuLysAlaIleGluAlaLeuSerGlyL 60
|||||
1 TGCCCGGACGAGCGT...GGCATCAAGGCCATCGAGCGCTTTCAGGTA 47

```
60 ysIleGluLeuHisGlyLysProIleGluValGluHisSerValProLys 76
|||||
48 AAATGGAACATACATGGGAAACCGATGGAAGTTGAGCACTCGGTCCCTAAA 97
|||||
77 ArgGlnArgIleArgLysLeuGlnIleArgAsnIleProProHisLeuGln 93
|||||
98 CGGCAGAGGATTCTGTAACCTTCAGATACGAATATCCCGCCCACTTACA 147
|||||
93 nTrpGluValLeuAspSerLeuLeuValGlnTyrGlyValValGluSerC 110
|||||
148 ATGGGAGGTGGTGATAGTTACTAGTCCAGTATGGAGTGGTGAGAGCT 197
|||||
110 ysGluGlnValAsnThrAspSerGluThrAlaValValAsnValThrTyr 126
|||||
198 CTGAGCAAGTGAACACGGATTTCGGAACGGCAGTTGTAAATGTAACCTAT 247
|||||
127 SerSerLysAspGlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGln 143
|||||
248 TCCAGTAGGACCAAGCTAGACAGCCCTAGACAAACTGAATGGATTCCA 297
|||||
143 nLeuGluAsnPheThrLeuLysValAlaTyrIleProAspGluMetAlaA 160
|||||
298 GTTAGAGAACTTCACCTGAAGTTGCCTACATCCACAGATGAAGTGTCTG 347
|||||
160 IaGlnGlnAsnProLeuGlnGlnProArgGlyArgArgGlyLeuGlyGln 176
|||||
348 CCCAGCAAAATCCCTCGGCACACAGCTCCGGGGCGCCGTGGGGCCAGGGCAG 397
|||||
177 ArgGlySerSerArgGlnGlySerProGlySerValSerLysGlnLysPr 193
|||||
398 CGGGGCTCATCCAGCAGCGCGTCTCCAGGATCGGTGTCCAGCAGAAACC 447
|||||
193 oCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGlyAlaI 210
|||||
448 CTGTGACCTGGCAGCTGGGCGCTGCTGTTCCACCAGTTTGTTCGAGCA 497
|||||
210 leIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226
|||||
498 TTATAGGAAAGAGAGTGTCTACCATTCGCAACATCACAAAGCAGACCCAG 547
|||||
227 SerLysIleAsp 230
|||||
548 TCTAGAATCGAT 559
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:23 ; Search time 5806.2 Seconds
(without alignments)
11879.492 Million cell updates/sec

Title: US-09-685-696-175
Perfect score: 4181
Sequence: 1 ggtggatgcgttggtgtgtt.....aaaaaaaaaaaaaaaaaaaaa 4181

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_on.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vi.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4165	99.6	4181	9	HSU097188	U97188 Homo sapien
2	4139	99.0	4155	9	HSU076705	U976705 Human putat
3	3994.2	95.5	4171	9	AF117108	AF117108 Homo sapi
4	3200	76.5	129432	9	AL023775	AL023775 Human DNA
c	5 2144.6	51.3	169739	9	AC005082	AC005082 Homo sapi
6	1511.2	36.1	1740	10	AB046173	AB046173 Mus muscu
c	7 1324	31.7	181369	2	AC092447	AC092447 Homo sapi
c	8 1324	31.7	188273	2	AC024199	AC024199 Homo sapi
9	1117	26.7	3147	5	AF042353	AF042353 Xenopus l
10	1116.4	26.7	2697	5	AF064633	AF064633 Xenopus l
11	1084.4	25.9	161404	9	AF002376	AF002376 Homo sapi
12	1084.2	25.9	2405	5	AF064634	AF064634 Xenopus l
13	1078.2	25.8	2161	5	AF055923	AF055923 Xenopus l
14	845.8	20.2	2011	5	AF161270	AF161270 Danio rer
c	15 842.8	20.2	158105	2	AC023375	AC023375 Homo sapi
16	820.4	19.6	2130	9	AF117106	AF117106 Homo sapi
17	817.2	19.5	2381	9	AF198254	AF198254 Homo sapi
18	814	19.5	2021	5	AF026527	AF026527 Gallus ga
19	802.8	19.2	2223	10	AF061569	AF061569 Mus muscu
20	710.4	17.0	2010	9	AF117107	AF117107 Homo sapi
21	695.6	16.6	3667	9	AF057352	AF057352 Homo sapi
22	658	15.7	2437	10	M05YBIC	L35549 Mus musculu
c	23 627.4	15.0	59747	2	AC069053	AC069053 Homo sapi
24	598.4	14.3	142971	9	AC020629	AC020629 Homo sapi
25	590.2	14.1	147556	2	AC011007	AC011007 Homo sapi
26	572.4	13.7	182695	2	AC015706	AC015706 Homo sapi
c	27 572.4	13.7	184653	2	AL596177	AL596177 Homo sapi
28	570.8	13.3	93682	2	AL139152	AL139152 Homo sapi
c	29 554	11.3	57261	2	AC090245	AC090245 Homo sapi
30	473.4	11.3	57261	2	AC090245	AC090245 Homo sapi
31	427	10.2	165600	2	AC079780	AC079780 Homo sapi
c	32 303.6	7.3	179501	2	AC090107	AC090107 Homo sapi
33	286.2	6.8	62635	2	AC087709	AC087709 Homo sapi
c	34 282	6.7	104668	9	AC021876	AC021876 Homo sapi
35	226	5.4	51119	2	AC090865	AC090865 Homo sapi
c	36 225.6	5.4	67823	2	AC026579	AC026579 Homo sapi
37	213.8	5.1	67823	2	AC026579	AC026579 Homo sapi
38	172.4	4.1	3188	3	AF241237	AF241237 Drosophill
39	136.8	3.3	158105	2	AC023375	AC023375 Homo sapi
40	122.4	2.9	258088	2	AC073766	AC073766 Mus muscu
41	100.2	2.4	64442	2	AC091133	AC091133 Homo sapi
c	42 100	2.4	159122	2	AC025556	AC025556 Homo sapi
c	43 95.4	2.3	201739	2	AC090288	AC090288 Mus muscu
c	44 94.8	2.3	239837	10	AC084407	AC084407 Mus Muscu
45	93	2.2	46580	2	AC012929	AC012929 Drosophili

ALIGNMENTS

RESULT	1	HSU097188	4181 bp	mRNA	PRI	11-SEP-1998
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						

Db 1501 CGGTGCCATCATCGCAAGAGGCGCAGCACATCAAGCAGCTTCTCGCTTTCGCTGGAGC 1560
QY 1561 ttcaattaaagattgtctcagcggaagcaccagatctaaagtgaagatggtgaattatcac 1620
Db 1561 TTCATTAAGATTGCTCCAGCGGAAGCACCAGATGCTAAAGTGAGATGGTGAATTATCAC 1620
QY 1621 tggaccaccagaggctcagttcaaggctcagggagaattttatggaaaaattaaagaaga 1680
Db 1621 TGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAATAATTAAGAAGA 1680
QY 1681 aaactttgttagtcttaagaagaaggtgaacttaagctcctataatcagagtgccatcett 1740
Db 1681 AAACATTAAGTCTTAAGAAAGAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTT 1740
QY 1741 tgcctgtgcagagttatttggaaaaagaggcaaaacggtgaatgaacttcagaatttgc 1800
Db 1741 TGCTGCTGCCAGAGTTATTTGGAAGAGGCAAAACGGTGAATGAACCTTCAGAAATTTGTC 1800
QY 1801 aagtgcaagaattgttgcctcgtgaccagacacactgatgagaatgaccaagtgtgtg 1860
Db 1801 AAGTGCAGAAAGTTGTTGCCCTCGTGACCAGACACCTGATGAGATGACCAAGTGGTTGT 1860
QY 1861 caaataaactggtcactctatgcttgcaggttgcagagaaataatcaggaattct 1920
Db 1861 CAAAATAACTGCTCACTTCTATGCTTGCCAGGTTGCCAGAGAAAATTCAGGAAATTCCT 1920
QY 1921 gactcaggttaagcagcaccacaacagagaagctctgaaagtgcacacactcagtcaag 1980
Db 1921 GACTCAGGTAAGCAGACCAACACAGAAAGGCTCTGCAAAAGTGACCACCTCAGTCAAG 1980
QY 1981 acggaagtaaggtcaggaacagccaccacagaggcagatgcacaaccaaagacaga 2040
Db 1981 ACGGAAGTAAAGGCTCAGGAACGCCACCACACAGAGGAGATGCCAACCAAGACAGA 2040
QY 2041 ttgcttaaccaacagatgggctgacccctatccagaaatcacatgcacaagtgtttac 2100
Db 2041 TTGCTTAACCAACAGATGGGCGCTGACCCCTATCCAGCAATCACATGCACAAGTTTTCAC 2100
QY 2101 ctagccaggtttctgaggaccaggaacttttgaactcctgctctgtgagaatgtat 2160
Db 2101 CTAGCCAGTTGTTCTTGAGGACCAGGCAACTTTTGAACCTCCTGCTCTGAGAAATGTAT 2160
QY 2161 actttatgctctctgaattgtatgacaccagcttttaaacaacaaacaaacaaacaa 2220
Db 2161 ACTTTATGCTCTCTGAAATGTATGACACCAGCTTTAAACAAACAAACAAACAA 2220
QY 2221 aaaaggttggggaggggaaagagaagagctctgcacttccctttgttagtctca 2280
Db 2221 AAAAGGTTGGGGAGGGGAAAGAGAGAGCTCTGCACCTTCCCTTTGCTGTAGTCTCA 2280
QY 2281 cagataacagatatcttaattctcttcttaattatcccccatatgccagaaattggctta 2340
Db 2281 CAGTATAACAGATATCTTAATTTCTTTTAATATCCCCCATATGCCAGAAATTTGGCTTA 2340
QY 2341 atgatgcttctcaaatcatcaaatagattgctctcaatccaatgttaaaattgga 2400
Db 2341 ATGATGCTTTTCACTAAATTCATAATAGATTGCTCTCTTAATCCCAATTTGTTAAATGGA 2400
QY 2401 tcagaataattatcagagaacttaaatgttaagccattagcatagaaaaactgttctca 2460
Db 2401 TCAGAATAATTATCAGAGAACTTAAATGTTAAGCCATTAGCATAGAAAAACTGTTCTCA 2460
QY 2461 gttttattttacctaactaacatagatgaacttaaggaagtgctgaatgggttggc 2520
Db 2461 GTTTTATTTTTTACCCTAACACTAACATGAGTAACCTTAAGGGAAGTGTGAATGGTGGC 2520
QY 2521 aggggtattaaacgctattttactcaactaccctcaggtattcagtaatacagaaga 2580
Db 2521 AGGGGTATTAAACGGTCATTTTACTTCAACTACCTCAGGTATTCAGTAATACAATGA 2580
QY 2581 gcaaaattgttcccttttttggaaaatttatatactttataatgatagaagtcacaacq 2640
Db 2581 AGTGTTTTCCTTAATGTTCTCTCTGAAAACAGTTTCTTTTGTAGTTTAAACCAAAAAGTG 2720

Db 2581 GCAAAATGTTCTCTTTTGTGAAAAATTTATATACTTTATATATAGATAGAAGTCCAACCG 2640
QY 2641 ttttttaaaaaataaattttaaatttaaacagcaactaacagcaactaacagcaactaacagcaact 2700
Db 2641 TTTTFTAAAAATAAATTTAAAAATTTAACAGCANTCAGTACAGGCAAAATTAAGATTTT 2700
QY 2701 tacttctggctggtagacagtaaaagctggaaaaataatttcagggttttttttgagggctttg 2760
Db 2701 TACTTCTGGCTGGTGCAGCTAAAGCTGGAAAAATTAATTTTCAAGGCTTTTGTGAGGCTTTG 2760
QY 2761 acacagttattagtttaaatcaaatgttcaaaaaataacggagcagtgctagtaactctgaga 2820
Db 2761 ACACAGTTATTAGTTAAATCAAAATGTTCAAAAATACGGAGCAGTGCCTAGTATCTGGAGA 2820
QY 2821 gcagcactaccattattcttccattatagttgggaagtttttgcaggttactacaaca 2880
Db 2821 GCAGCAGTACCATTATTTCTTCAATTTATAGTTGGGAAGTTTTCACGCTACTAACAAA 2880
QY 2881 gtggctgcaggagattttggaaaggctgggttttaattgggttcaggagacttcagttttt 2940
Db 2881 GTGGTGCAGGAGATTTTGGAACGGCTGGTTTAAATGGCTTTCAGGAGACTTCAGTTT 2940
QY 2941 gtttagctacatgattgaatgataataaatgctttgtctctgactatcaataactaa 3000
Db 2941 GTTTAGCTACATGATTGAATGCATAATAATGCTTTGTGCTCTGAATATCAATACCTAA 3000
QY 3001 aqaaagtcagtcagtaagagatgcaagctttcaactgactggcgaagaagcaagcttta 3060
Db 3001 AGAAAGTGCATCAGTGAAGAGATGCAAGACTTTCAACTGACTGGCAAAAAGCAAGCTTTA 3060
QY 3061 gctgtcttataggatgcttagtttggccactacacttcagaccaatgggacagctcataga 3120
Db 3061 GCTTGTCTTATAGGATGCTTAGTTTGCCTACTACACTTCAGACCAATGGGACACTCATAGA 3120
QY 3121 tgggtgacagtggtttaaagcgaacaaaggctacatttccatggggcgcagcactgtcat 3180
Db 3121 TGGTGTGACAGTGTTTAAACGCCAACAAAGGCTTACATTTCCATGGGCCACGACTGTCAT 3180
QY 3181 gggcctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaa 3240
Db 3181 GAGCCTCATTAGCTATTTTGAAGATTTTAAAGCCTGATAAATTTAAAAAATAAAAAA 3240
QY 3241 aaattagactccacttaagtagtaaaagataaagcagatttctgtactctgtgtcaatcag 3300
Db 3241 AAATTAGACTCCACCTTAAGTAGTAAAGTATAACAGGATTTCTGTATCTGTGCAATCAG 3300
QY 3301 ttctttgaaaaaaatgcataaagataagagaatacacagaaaaagtttngggatataattg 3360
Db 3301 TTCTTTGAAAAAAAAGTCAAAAGATAGAGAAATACAGAAAAGTTTNGGGATATAAATTTG 3360
QY 3361 aatgactgtgaaaaacatatgacctttgataacgaactcatttgcactccttgacagca 3420
Db 3361 AATGACTGTGAAAAACATATGACCTTTGATAACGAACCTCATTTGCTCACTCCTTGACAGCA 3420
QY 3421 aagccagtcagtcacaaattgtgtgggtggtgggtggttcccaaggccagcgtctctctg 3480
Db 3421 AAGCCAGTACGTACAAATTTGTTGGTGGTGGTGGTCTCAAGGCCACGCTGCTCTCTG 3480
QY 3481 aattgatttttggagtttggnttgnaaagatgatacagncatgtttacactgatcttnaa 3540
Db 3481 AATTGATTTTTTGGTTTTGGNTTGNAAAGATGATACAGNCATGTTACACTGATCTTNA 3540
QY 3541 ggaatatattataaccccttttaaaaaataatccccctgctcacttcttattttcgagatga 3600
Db 3541 GGACATATTNTTAAACCCCTTTAAAAAAAATCCCTGCTCAATCTTATTTATTCGAGATGA 3600
QY 3601 atttcagatcagactagatgtcttctgaagatcaattagacattntgaaaaatgatttaa 3660
Db 3601 ATTTCGATACAGACTAGATGTCTTTCTGAAGATCAATAGACATTTGTGAAAAATGATTTAA 3660
QY 3661 agtggtttcccttaagtctctgaaaaacaaagttctttttagtttttaaccacaaagtg 3720
Db 3661 AGTGTTTTTCCTTAATGTTCTCTCTGAAAACAGTTTCTTTTGTAGTTTAAACCAAAAAGTG 3720

Db 901 CATTCGGAACATCACAAACAGACCAGCTCAAAATCGATGTCCACCGTAAGAAAAATGCG 960
 Qy 961 999ggtgctgagaagtgattactatctcttactctctgaagcgaacctctgcggttg 1020
 Db 961 GGGGCTCTGAGAAGTCGATTACTATCTCTTACTCTCTGAAGGCACCTCTGCGGCTTG 1020
 Qy 1021 taagtctattctgagagattatgcataaagagctcaagatatataaattccagaaagagat 1080
 Db 1021 TAAGTCTATTCTGGAGATTATGCTATAAGAAAGCTCAAGATATAAATTCACAGAAAGAT 1080
 Qy 1081 cccctgaagattttagctcataataactttgttgagcgtcttattgtaagaaggag 1140
 Db 1081 CCCCTGAAGATTTAGTCTATAAATCTTGTGGAGCTCTTATGTTGAAGAAGGAAG 1140
 Qy 1141 aaatctaaaaaattgagcaagcacagacactaaaatcacgatctcattcattgcagga 1200
 Db 1141 AAATCTTAAAAAATTGAGCAAGACACAGACATPAAAAATCAGATATCTCCATTCACAGA 1200
 Qy 1201 attgacgctgataatccagacgcactattacaggttaaagcaatgttgagacatgtgc 1260
 Db 1201 ATTGACGCTGATATATCCAGAACGCATATTACAGTTAAAGCAATGTTGAGACATGTGC 1260
 Qy 1261 caaagctgagagagatcatgaagaaaaatcaggagctcttatgaaaaatgatattgcttc 1320
 Db 1261 CAAAGCTGAGGAGGAGATCATGAAGAAATCAGGAGGCTTATGAAAAATGATATTGCTTC 1320
 Qy 1321 tatgaatctcaagcacatttaactcctggattaaatctgaacgcttggctgtgtccc 1380
 Db 1321 TATGATCTTCAAGCACATTTAATTCCTTGGATTAATCTGAACGCCCTTGGGCTGTCTCCC 1380
 Qy 1381 acccaactcagggatgccacctccacactcagggcccttcagccatgaactccctcccta 1440
 Db 1381 ACCCACTTCAGGGATGCCACCTCCACCTCAGGGCCCCCTTCAGCCATGACTCCTCCCTA 1440
 Qy 1441 cccgcagtttgaagaatcagaacaggaactgttcaatcagtttatccacagctctatcagtt 1500
 Db 1441 CCCGCAGTTTGAGCAATCAGAAACGGAGACTGTTCAATCAGTTTATCCAGACTCTATCAGT 1500
 Qy 1501 cgggtccatctcggcagcaggccagcacatcaagcagcttctcgtcttctggagc 1560
 Db 1501 CGGTCCCATCATCGGCAAGAGGCCAGCACATCAAGCAGCTTCTCGCTTTGCTGGAGC 1560
 Qy 1561 ttcaattaagattgtccagcggagcaccagatgtcctaagtggagtgattatcac 1620
 Db 1561 TTCAATTAAATGCTCCAGCGAAGCACCAGATGCTTAAGTGAGGATGGTGATTATCAC 1620
 Qy 1621 tggaccaccagagctcagttcaaggctcagggaagaatttatggaaaaattaaagaaga 1680
 Db 1621 TGGACCACAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAATTAAGAAGA 1680
 Qy 1681 aaactttgttagtctaagaagagtgaaacttgaagctcatatcagagtgcaatcctt 1740
 Db 1681 AAACTTTGTAGTCTTAAAGAAGAGGTGAACCTTGAAGCTATATFCAGAGTGCCATCCTT 1740
 Qy 1741 tgcgtcggcagagttatttgaaaaggaggcaaacggtgaaatccagaaattgtc 1800
 Db 1741 TGCTGCTGCAGAGTTATTGGAAAGGAGGCAAAACGGTGAATGAACCTTCAGAATTTGTC 1800
 Qy 1801 aagtcagaagtgtgttgcctctgtgaccagacacacctgatgagaatgaccgaagtgtgtg 1860
 Db 1801 AAGTGCAGAAGTTGTTGTCCCTCGTGACCAGACACCTGATGAGATGACCAAGTGTTGT 1860
 Qy 1861 caaaaactggtcacttctatgttgcaggttggccagagaaaaattcaggaaattct 1920
 Db 1861 CAAAATAACTGGTCACTTCTATGCTTGGCAGGTGCCCCAGAGAAAAATTCAGGAAATTC 1920
 Qy 1921 gactcaggttaagcagcaccacaacagaggtctctgaaagtggaacacctcagtcag 1980
 Db 1921 GACTCAGGTAAAGCAGCACCAACACAGAGGCTCTGCAAGTGGACCACTCAGTCAAG 1980
 Qy 1981 acggaagtaaaggctcaggaacagccaccacagagggcagatgccaacccaaagacaga 2040
 Db 1981 GCTTGTCTTATAGGATGCTTAGTTGGCCACTACACTTCAGACCAATGGGACAGTCATAGA 3120

Db 1981 ACGGAAGTAAAGGCTCAGGAACAGCCACCACACAGAGGCGAGATGCCAAACCAAGACAGA 2040
 Qy 2041 ttgcttaacaaacagatggcgctgaccccttatccagaatccacatgcacaagtgttttac 2100
 Db 2041 TTGCTTAACCAACAGATGGCGCTGACCCCTTATCCAGAATCACATGCACAAGTTTTTAC 2100
 Qy 2101 ctagccagttgtttctgaggaaccaggcaactttgaactcctgtctctgtgagaatgtat 2160
 Db 2101 CTAGCCAGTTGTTCTGAGGACAGGCAACTTTTGAACCTCTCTCTCTGTGGAATGTAT 2160
 Qy 2161 actttatctctcgtaaaatgtatgacaccagcttttaaacaacaaacaaacaaacaaa 2220
 Db 2161 ACTTTATGCTCTCTGAATGTATGACACCCAGCTTTAAACAAACAAACAAACAAACAA 2220
 Qy 2221 aaaaagggtggggagggaggaagagagctctgcacttccctttgttagtctca 2280
 Db 2221 AAAAGGTTGGGAGGAGGAGAAAGAGAGCTCTGCACCTCCCTTTCTGTGTAGTCTCA 2280
 Qy 2281 cagtaaacagatatcttaattcttcttaatttcccccaataatgccagaaaaattggtcta 2340
 Db 2281 CAGTATAACAGATATTCTAAATCTTCTTAAATTCCTCCCATTAATGCCAGAAATTTGCTTA 2340
 Qy 2341 atgatgctttcaactaaaattcatcaaatagattgctccttaaatcccaattgttaaaattgga 2400
 Db 2341 ATGATGCTTTCACTAAATTCATCAATAGATTGCTCTAAATCCCAATTTGTTAAATTTGA 2400
 Qy 2401 tcagaataattatcacaggaacttaattgaagcatttagcatagaaaaactgttctca 2460
 Db 2401 TCAGAAATTAATATCAGGAACTTAAATGTTAAGCCATTTAGCATAGAAAACTGTTCTCA 2460
 Qy 2461 gttttatttttaccataacactaactgaaggaagtgctgaatgggttggtgc 2520
 Db 2461 GTTTTATTATTTACCTAAACACTAACATGAGTAACCTAAGGGAAGTGTGAATGGTGTGC 2520
 Qy 2521 aggggtattaaacgtgcatttttactcaactcactcagggtattcagtaataacaaagaaa 2580
 Db 2521 AGGGGTATTAAACGTGCATTTTACTCAACTACCTCAGGTATTCAGTAATACAAATGAAA 2580
 Qy 2581 gcaaaatgttcccttttttgaaaatttatatactttataatgatagaagcccacacg 2640
 Db 2581 GCAAAATTTGTTCCCTTTTTCGAAAAATTTATATACTTTATATGATAGAAAGTCCAAACG 2640
 Qy 2641 ttttttaaaaaataaatttaaatttaacagcaactcagctaaacagcagcaaatlaagatttt 2700
 Db 2641 TTTTTTAAAAATAAATTTAAATTTAACAGCAATCAGCTAACAGGCAAAATTAAGATTTT 2700
 Qy 2701 tactctcgtcgtgagcagtaaaagctggaataatlaatttcagggttttttggaggttttg 2760
 Db 2701 TACTTCTGGCTGGTGACAGTAAGCTGGAAAAATTAATTTTCAGGGTTTTTTTGAGGCTTTTG 2760
 Qy 2761 acacagttattagttaaatcaaatgttcaaaaatcaggaagcagtgctagtatctgaga 2820
 Db 2761 ACACAGTTATTATAGTTAAATCAAAATGTTCAAAAATACGGAGCAGTGCTTAGTACTCGAGA 2820
 Qy 2821 ggcagcactaccatttattcttttcattttagttggaaaagtttttgcggttaactaaacaa 2880
 Db 2821 GCAGCATTACCATTTATTTCTTTCAITTTATAGTTGGGAAGTTTTCACGGTACTACAAA 2880
 Qy 2881 gtggtcgcagagagatttggaacgctggtttaaatggttcaggagagactcagtttttt 2940
 Db 2881 GTGTGTCGAGGAGATTTTGGAAACGGCTGGTTTAAATGGCTTCAGGAGACTTCAGTTTTTT 2940
 Qy 2941 gtttagctacatgattgaatgataataaattgttgccttctgaactatcaataactaa 3000
 Db 2941 GTTTAGCTACATGATTGAATGCATATAAATGCTTTGTGCTCTTGACTATCAATACCTAA 3000
 Qy 3001 agaaagtgcacagtgaaagagatgcaagactttcaactgactggcgaagaaagcaagcttta 3060
 Db 3001 AGAAAGTCATCAGTCAAGAGATGCAAGACTTTCAACTGACTGGCAAAAAAGCAAGCTTTA 3060
 Qy 3061 gcttcttattaggatgcttagttgcccactacacttcagaccaaattggaagagtcataga 3120
 Db 3061 GCTTGTCTTATAGGATGCTTAGTTTGGCCACTACACTTCAGACCAATGGGACAGTCATAGA 3120

|||||
Db 197 TCCATACCGTGGGTGTTTAAATTTTCGTTTTCCTCGTTATTTTAAACAAACC 256
Qy 241 actctcacaaatgaacaaactgtatatcggaacctcagcgagaaacgcccgcctcgga 300
|||||
Db 257 ACTCTTCACAATGAACAACTGTATATCGGAACCTCAGCGAAGACGCGCCCTCGGA 316
Qy 301 cctagaagtattctcaagagacccaaga tcccggtgctcgggacctctctgtggaagac 360
Db 317 CCTAGAAGTATCTTCAAGACGCGCAAGATCCCGGTGTCGGGACCTTCTCGTGAAGAC 376
Qy 361 tggctacgcttcgtgactgcccgaacgagagctggccctcaaggccatcgaggcgt 420
|||||
Db 377 TGGCTACCGGTGCTGACATGCGCGGACGAGAGTGGCCCTCAAGGCCATCGAGGCGCT 436
Qy 421 ttcaggTaaatagaactgcacgggaacccatagaagttgagcactcggTcccaaaaag 480
Db 437 TTCAGGTAAATAGAAGTGCACGGGAACCCATAGAAAGTTGAGCACTCGGTCCCAAAAG 496
Qy 481 gcaagattcggaaacttcagatacgaataatccccctcatttacagtgggaggct 540
Db 497 GCAAGGATTTCGGAACCTTCAGATACGAATATCCCGCCTCATTTACAGTGGGAGGTGCT 556
Qy 541 ggatagttactagtcaggtatggagtgggtggagagctgtgagcaagtgaacactgactc 600
Db 557 GGATAGTPTTACTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 616
Qy 601 ggaactgcagttgtaagtgaacttaccattccagtaagcaacaaagctagacaagcactaga 660
Db 617 GGAACGTGACAGTTGTAATGTAACTTATTCAGTAAAGACCAAGCTAGAACACACTAGA 676
Qy 661 caaactgaattggatttcagttagagaatttcacottgaaagtagcttatccccctgatga 720
Db 677 CAAACTGAATGGATTTCAGTTAGAGAAATTCACCTTGAAAGTAGCCTATATCCCTGATGA 736
Qy 721 aatggccgccacgaaccccccttgagagcccccgggtggccgggtggccggag 780
Db 737 AATGGCCGCCACGAACCCCTTGCAGCAGCCCGGAGTGGCGGGGGTGGCGCAGAG 796
Qy 781 gggctctcaagcaggggtctccaggtccgtatccaaagcaaaacacatgtattgccc 840
Db 797 GGGCTCTCAAGCAGGGGTCTCCAGAGTCCGTATCCAGCAAGAAACATGTGATTGCCC 856
Qy 841 tctgcgctgctggttccccaccaaattgttgagccatcatagaaagaaggtgccac 900
Db 857 TCTGCGCTGCTGGTTCCTCCACCAATTTGTTGGAGCCATCATAGAAAGAGGTGCCAC 916
Qy 901 cattcgaaacatcaccaaacagaccagctctaaatcgatgtccaccgttaagaaaaatgc 960
Db 917 CATTCGGAACATCACCAAAACAGACCCAGTCTAAATCGATGTCCACCGTAAAGAAAAATGC 976
Qy 961 gggggctgctgagaagtcgattactatccctctactctcctgaaggcacctctcgcgcttg 1020
Db 977 GGGGGCTGCTGAGAAAGTCGATTAATCTCTCTACTCTCTGAAAGCCACTCTCGCGGCTTG 1036
Qy 1021 taagtctattctggagattatgataaaggagctcaagatatataaattcaagaagagat 1080
Db 1037 TAAGTCTATTCTGGAGATTATGCATAAGGAAGCTCAAGATATAAATTCACAGAAGAGAT 1096
Qy 1081 ccccttgaagattttagctcataataacttgttggacgttcttatttggtaagaaggaag 1140
Db 1097 CCCCCTGAAGATTTTAGTCTATATAACTTTGTTGGACGCTTATTGTTGAAGGAAGAAAG 1156
Qy 1141 aaactctaaaaaattgagcaaacacagacactaaatcacgatattccatttcagga 1200
Db 1157 AAATCTTAAAAAATTGAGCAAGACACAGACACTAAATACGATATCTCCATTTCAGGA 1216
Qy 1201 attgacgctgtataatccgaacacacactattacagtttaaaggcaatgtgagacatgtgc 1260
Db 1217 ATTGACGCTGTATAATCCAGACGCACTATTACAGTTAAAGGCAATGTTGAGACATGTGC 1276
Qy 1261 caaagctgaggagagatcatgaagaaaaatcaggagctcttatgaaaaatgatattgttc 1320
|||||

Db 1277 CAAAGCTGAGAGGAGATCATGAAGAAATCAGGGAGTCTTATGAAATGATATTGTCTTC 1336
Qy 1321 tatgaatttcaagcacatttaattccctggattaaatctgaacgcttgggtctgttccc 1380
|||||
Db 1337 TATGAATCTTCAAGCACATTTAATTCCTGGATTAAATCTGAACGCTTGGGTCTGTCTCC 1396
Qy 1381 accacttcagggtatgcacactccacactcaggggcccccttcagccatgaactccctccta 1440
Db 1397 ACCCACTTCAGGGATGCCCACCTCCACCTCAGGGCCCCCTTCAGCCATGACTCTCCCTA 1456
Qy 1441 ccccgagtttgacaatcagaaacgagactgttcactcagtttatcccgactctcatcagt 1500
Db 1457 CCCGAGTTTGAGCAATCAGAAACGAGACTGTTCTATCTGTTTATCCAGCTCTATACAGT 1516
Qy 1501 cggTgcctatcatcggaagcagggccagcacatcaagcagctttctcgttttgctggagc 1560
Db 1517 CGGTGCCATCATCGCAAGCAGGGCCAGCACATCAAGCAGCTTCTCGCTTTGCTGGAGC 1576
Qy 1561 ttcataaagattgctccagcgggaagcaccagatgtctaagtgaagtagtgatattacac 1620
Db 1577 TTCAATTAAAGATTGCTCCAGCAGAAAGCACCAGATGCTAAAGTGAGGATGGTGAATATCAC 1636
Qy 1621 tggaccaccagagctcagttcaaggctcaggggaagaatttatgaaaaattaaagaaga 1680
Db 1637 TGGACCCACGAGGCTCAGTTCAAGGCTCAGGGAAGAAATTTATGAAAAATTTAAAAAGA 1696
Qy 1681 aaactttgttagtctaaagaagaggtgaacttgaagctcatatcagagtccatcctt 1740
Db 1697 AAACCTTGTAGTCTTAAAGAAAGAGGTGAACCTTGAAGCTCATATCAGAGTCCCATCTT 1756
Qy 1741 tgcTgctggcagagttatttgaaaaaggaggcaaacggtgaatgaacttcagaatttgc 1800
Db 1757 TGCTGCTGGCAGAGTTATTGGAAAAAGGAGGCAAAACGGTGAATGAATTCAGAATTTGTC 1816
Qy 1801 aagtgcagaagtgtgtccctcgtgaccagacactgatgagaatgaccagaagtgtgt 1860
Db 1817 AAGTGCAGAAATGTTGTCCCTCGTGACACAGACCTGATGAGATGACCAAGTGGTGT 1876
Qy 1861 caaaactcagtgctcacttctatgcttgcaggttggccagagaaaaattccaggaattct 1920
Db 1877 CAAAATACTGGTCACTTCTATGTTGCCAGGTTCGCCAGAGAAATTCAGAGAAATCT 1936
Qy 1921 gactcaggtaaagcagcaccacaaacagaaggtctgcaaaagtgaccacactcagtcag 1980
Db 1937 GACTCAGTTAAAGCAGCACCAACAAACAGAGGCTCTGCAAAAGTGACCACTCAGTCAAG 1996
Qy 1981 acggaagtaaggctcaggaacacagccacacagagcgagatgccaaacaaagacaga 2040
Db 1997 ACGGAAGTAAAGGCTCAGGAACACGCCACCAAGAGGAGATGCCAACCAAGACAGA 2056
Qy 2041 ttgcttaaccacagatggcgctgaaccccttatccagaatcacatgcacagttttac 2100
Db 2057 TTGCTTAACCAACAGATGGGCGCTGACCCCTTATCCAGAAATCAGATGACAAATTTTAC 2116
Qy 2101 ctagccagttgtttctgaggaccaggcaacttttgaactcctgctctctgtgagaatgtat 2160
Db 2117 CTAGCCAGTGTCTGTAGGACCAAGCAACTTTTGAACCTCTCTCTGTGTGAGAAATGTAT 2176
Qy 2161 actttatgctctgaaaatgtatgacccccagctttaaaaaacaaacaaacaaacaaa 2220
Db 2177 ACTTTATGCTCTCTGAAATGTATGACACCCAGCTTTTAAACAAACAAACAAACAAA 2236
Qy 2221 aaaaggttggggaggagggaagagagagctctgcacttcccttctgtttagtctca 2280
Db 2237 AAAAGGTGGGGAGGAGGGAAGAGAGAGCTCTGCACITCCCTTTGTTGTAGTCTCA 2296
Qy 2281 cagtataacagatatcttaattcttcttaatttcccccaataatccagaaaatttggctta 2340
Db 2297 CAGTATAACAGATATTCTAATTTCTTAAATATTCCTCCCATTAATGCCAGAAATTTGGCTTA 2356
Qy 2341 atgatcttccactaaattcatcaaatagattgtctctaaatcccaattgttaaaattgga 2400
Db 2357 ATGATGCTTTTCACTAAATTCATAAATAGATTGCTTCTTAAATCCAAATTTGTTAAAAATTGGA 2416
|||||

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 497J21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 427A4 (298049) is at 129333 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

Assembly was confirmed by BamHI and HindIII digests. Discrepancy noted in EcoRI digest.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

497J21 is from the library RPC13 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see [http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2](http://bacpac.med.buffalo.edu/VECTOR: pcYPAC2).

FEATURES

	source
	1. 129432
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/map="q26-27"
	/clone="RP3-497J21"
	/clone_lib="RPC1-3"
	complement(1..62)
repeat_region	/note="AluX repeat: matches 62. .1 of consensus"
misc_feature	27..424
repeat_region	/note="match: STS AL022447"
repeat_region	63..231
repeat_region	/note="MER20 repeat: matches 53. .218 of consensus"
repeat_region	433..591
repeat_region	/note="3 copies 53 mer 86% conserved"
repeat_region	complement(1513..1915)
repeat_region	/note="MLT1B repeat: matches 382. .2 of consensus"
repeat_region	complement(2671..2713)
repeat_region	/note="MLT1B repeat: matches 47. .3 of consensus"
repeat_region	complement(3291..3586)
repeat_region	/note="AluSq repeat: matches 298. .1 of consensus"
repeat_region	4989..5128
repeat_region	/note="L1MB6 repeat: matches 1. .141 of consensus"
repeat_region	5137..5648
repeat_region	/note="L1MB5 repeat: matches 382. .923 of consensus"
repeat_region	complement(7362..7463)
repeat_region	/note="MIR repeat: matches 194. .87 of consensus"
repeat_region	8372..8668
repeat_region	/note="AluX repeat: matches 2. .300 of consensus"
repeat_region	8670..8703
repeat_region	/note="17 copies 2 mer aa 82% conserved"
repeat_region	14252..14295
repeat_region	/note="22 copies 2 mer gt 98% conserved"
repeat_region	14254..14301
repeat_region	/note="3 copies 16 mer 94% conserved"
repeat_region	14254..14297
repeat_region	/note="11 copies 4 mer gtgt 98% conserved"
repeat_region	14441..14481
repeat_region	/note="MIR2 repeat: matches 99. .139 of consensus"
repeat_region	complement(16259..16324)
repeat_region	/note="MIR repeat: matches 157. .92 of consensus"
repeat_region	complement(18532..18711)
repeat_region	/note="MIR repeat: matches 218. .29 of consensus"
repeat_region	20727..21002
repeat_region	/note="AluX repeat: matches 2. .296 of consensus"
repeat_region	22965..23004
repeat_region	/note="10 copies 4 mer caca 83% conserved"
repeat_region	complement(24989..25171)
	/note="L1MB3 repeat: matches 931. .738 of consensus"
	complement(25204..25504)
	/note="AluSp repeat: matches 303. .1 of consensus"
	complement(25761..26017)
	/note="AluY repeat: matches 301. .45 of consensus"
	complement(27102..27194)
	/note="MIR repeat: matches 142. .50 of consensus"
	30090..30204
	/note="5 copies 23 mer 94% conserved"
	complement(31383..31419)
	/note="MIR2 repeat: matches 113. .76 of consensus"
	31912..32368
	/note="match: STS S53188; part 1 of genomic marker D6S193"
	32276..33267
	/note="31 copies 32 mer 91% conserved"
	33141
	/note="tandem repeat"
	33204
	/note="tandem repeat"
	33222..33396
	/note="tandem repeat and single clone region"
	33268..33555
	/note="9 copies 32 mer 88% conserved"
	33288..33539
	/note="17 copies 16 mer 89% conserved"
	33473..33793
	/note="match: STS S53190; part 2 of genomic marker D6S193"
	complement(34751..35259)
	/note="MER21B repeat: matches 754. .240 of consensus"
	complement(35356..35625)
	/note="MER21B repeat: matches 282. .5 of consensus"
	36446..36618
	/note="MLT1B repeat: matches 1. .179 of consensus"
	36619..36921
	/note="AluSq repeat: matches 1. .303 of consensus"
	36927..37136
	/note="MLT1B repeat: matches 175. .390 of consensus"
	38179..38310
	/note="2 copies 66 mer 92% conserved"
	38187..38318
	/note="4 copies 33 mer 91% conserved"
	complement(39918..40225)
	/note="AluX repeat: matches 302. .1 of consensus"
	complement(41241..41409)
	/note="L1 repeat: matches 5390. .5208 of consensus"
	42029..42091
	/note="MIR2 repeat: matches 69. .138 of consensus"
	complement(43307..43755)
	/note="match: GSS B76271"
	46203..46713
	/note="MER1A repeat: matches 1. .527 of consensus"
	complement(50808..51108)
	/note="AluSp repeat: matches 303. .2 of consensus"
	52433..52723
	/note="AluDb repeat: matches 5. .301 of consensus"
	53862..54166
	/note="MER2 repeat: matches 1. .324 of consensus"
	54319..54430
	/note="7 copies 16 mer 81% conserved"
	54322..54417
	/note="3 copies 32 mer 89% conserved"
	54720..54767
	/note="24 copies 2 mer tg 88% conserved"
	58071..58370
	/note="AluX repeat: matches 1. .301 of consensus"
	58587..58738
	/note="4 copies 38 mer 86% conserved"
	complement(59963..60560)
	/note="match: GSS AQ001508"
	60205..60273
	/note="3 copies 23 mer 84% conserved"
	61253..61299
	/note="MIR repeat: matches 104. .150 of consensus"

Qy 3325 tagaataacaagaaagtgttngggatataatttgatgactgtgaaacatatgacct 3384
|||||
Db 147560 TAGAATAACAAGAAAGTGTGCGATATAATTGAATGACTGTGAAACATATGACCT 147501

Qy 3385 ttgataacgaactcatttgcctcaactccttgacagcaagaccagctacacaaattgtctt 3444
|||||
Db 147500 TTGATAACGAACCTCATTTGGCTACCTCTGACAGCAAGACCCAGTACGATCAAAATGTGTT 147441

Qy 3445 ggggtggtggtggtctcaagggccagctgctctcgaattgatttttggatttggntt 3504
|||||
Db 147440 GGGTGTGGTGTCTCCAAGCCAGCTGCTCTCTGAATTGATTTTGTGAGTTT-GTTT 147382

Qy 3505 gnaagatgatcacagncatgtttacactgatctttnaaggacatatntataaccctttaaa 3564
|||||
Db 147381 GTAAGATGATCACAGTCATGTATACACTGATCTAAAGG---ACATATATATAACCCCTTTAA 147325

Qy 3565 aaaaaaacccctccctcattcttatttcgagatgaatttcgatacagactagatgtctt 3624
|||||
Db 147324 AAAAAAATCACTGCCCTCATCTTATTTCAAGATGAATTTCTATACAGACTAGATGTTT 147265

Qy 3625 tctgaagatcaattagacattntgaagaatgatttaagtggttttcccttaattgttcttga 3684
|||||
Db 147264 TCTGAAGATCAATTAGACATTTTGAAGAATGATTTAAAGTGTTTCTCTTAATGTTCTCTGA 147205

Qy 3685 aacaagttcttctttagttttaacccaaaaagtgcccttttctgactggtttctctt 3744
|||||
Db 147204 AACAAGTTTCTTTGTAGTTTAAACCAAAAAGTGCCCTTTTGTCACTGGATTCCT 147145

Qy 3745 agcatcatgatttttttccacacaaatgaattaaaaattgtctaaatcatggaactggctt 3804
|||||
Db 147144 AGCATTCATGATTTTCTTTCATACAAATGAATTTAAATTTGCTAAATCATGGAAGTGGCTT 147085

Qy 3805 tctggttgatttcagtaaatggtttaaaggccagagcttttctcagatttgatttt 3864
|||||
Db 147084 TCTGTTGGATTTTCAAGTAAGATGTTTAAAGCCAGAGCTTTTCTCAGTATTGATTTT 147025

Qy 3865 ttccccaattattgattttttaaaaaatacacatagagctgcatctaaacacctgctg 3924
|||||
Db 147024 TTTCCTCAATATTGATTTTAAATAATACACATAGTGTCTGCATTTAT-ATCTGCTG 146966

Qy 3925 gtttaattctgtcanatttccactcttagccttttagtatgctgaatcanaanattacttt 3984
|||||
Db 146965 GTTTAAATTCGTCTATATTTTCTTCTAGCCTTTTAGTATGCAAAATCATATTTTACTTT 146906

Qy 3985 tactaagcatttgaatttgagtatctggtactagctgaagaataattcnaatttga 4044
|||||
Db 146905 TACTTAAGCATTTGTAATTTGGAGTATCTGCTAGTACTAGCTAAGAAATAATTTCTATAATTGA 146846

Qy 4045 gtttgtactcncaaaanattgggtcattctcctcattgnataatgtntcccccaatgcagcttc 4104
|||||
Db 146845 GTTTTGACTCACCATATATGATGATTCCTTCATGTATGATAATGTGCCCCAATGAGCTTC 146786

Qy 4105 atttccaganacctgacgcagataaaatttttcatcatttaggtcc 4153
|||||
Db 146785 ATTTTCCAGATACCTTGACGAGAATAAATTTTTCATCATTTAGGTGC 146737

RESULT 6
AB046173 1740 bp mRNA ROD 23-DEC-2000
LOCUS Mus musculus mimp3 mRNA for igf2 mRNA-binding protein 3, complete cds.
DEFINITION
AB046173
AB046173.1 GI:11933383
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognatha; Muridae; Mus.
REFERENCE
1 (sites)
Mori,H., Sakakibara,S., Imai,T., Nakamura,Y., Iijima,T., Suzuki,A.,
Yuasa,Y., Takeda,M. and Okano,H.

TITLE Expression of mouse igf2 mRNA-binding protein 3 and its implications for the developing central nervous system
J. Neurosci. Res. (2001) In press
REFERENCE 2. (bases 1 to 1740)
AUTHORS Mori,H., Okano,H., Sakakibara,S. and Imai,T.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) Hiroshi Mori, Osaka Univ. Graduate School of Medicine, Division of Neuroanatomy (D12); 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:hmori@nana.med.osaka-u.ac.jp, Tel:81-6-6879-3581, Fax:81-6-6879-3589)

FEATURES
source
1..1740
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..1740
/gene="mimp3"
CDS
1..1740
/gene="mimp3"
/codon_start=1
/product="igf2 mRNA-binding protein 3"
/protein_id="BAB19755.1"
/db_xref="GI:11933384"
/translation="MNKLYIGNLSDHAGPADLESVFKDAKIPVAGPFLVKTGYAFVDC
PDEGWLKAI EALSGKMEHGRKMEVHSVPKRQRIRKQIRNIPPHLQEWELDSLLY
QYGVVSECEQVNTDSEYAVNVYTSKDOARQALDKLNGFQLENTLVKWAYIPDETAA
QONSPQLRRGRGPGORSSQAPSGVSKQPCDLPLRLPLVPTQFVGAIIQKGTAEI
RNITKQTSKIDVHRKENTGAEKSIITLSTPEGTSAAKSLTEIMHKEADQIKFTEE
IPLKILAHNNFVGRLLIGKEGRNLKIEQDQDTKTIISPLQELTLNPERTIVKGSVE
TCAKEEIMMKIRIESYENEDIASMNLQHLIPGLNLNALGLFFPSGMPPTSGGPST
LTPYQFQESQETETVHLFIPALSVGAIIQKQHQHILKLSRFAGISIKIAPAEAPDAK
VWMIITITGPPAQKQGIYKIKKEENFVSPKEEVKLEAHRVPSFAAGRVIGKGGK
TVNELQSLSAEAVVVPDQTPDENQVVVKITGHFYACQVAKRQIEILTVQKQHQQ
KALOSGPPQSRKK"

BASE COUNT 500 a 431 c 442 g 367 t
ORIGIN
Query Match 36.1%; Score 1511.2; DB 10; Length 1740;
Best Local Similarity 91.8%; Pred No. 1.3e-294;
Matches 1597; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 251 atgaacaactgtatatcggaacacctcagagaaagccgcgcctcgcacccctcgacccctagaaagt 310
|||||
Db 1 ATGAACAAATTTGTACATCGGAACCTCAGCGACCAGCGCGGCCGCCGCGACCTGGAAAGT 60

Qy 311 atcttcaaggacgcgaagatcccggtgtcgggacccttctcgttgaagactggctacgag 370
|||||
Db 61 GTCTTCAAGGACGCTAAGATCCCGGTGGCGGCCCTTCTCTGTTGAAGACGGGCTACGG 120

Qy 371 ttcgttgactgcccgagagagctgggcccctcaaggccatcgagcgccttcaggtaaa 430
|||||
Db 121 TTCGTGGACTGCCCGGACGAGGGCTGGGGCCCTCAAGGCCATCGAGCGCTTCAGGTAAA 180

Qy 431 atagaactgcacgggaaccccatagaagttagcactcgggtcccaaaaggcaaaagatt 490
|||||
Db 181 ATGGAATACATGGAACCCGATGGAAGTTGAGCACTTCGTCCTTAACCGCAGAGGATT 240

Qy 491 cggaaacttcagatacgaataatccgcctcatttacagtgggagtgctggaatagttta 550
|||||
Db 241 CGTAAACTTCAGATACGAATAATCCCGGCCCACTTACAATGGGAGGTGCTGGATAGTTTA 300

Qy 551 ctatgccagtatggagtggtgagagctgtgagcaagtgaacactgactcggaaactgca 610
|||||
Db 301 CTAGTCCAGTATGGAGTGGTGAGAGCTGTGACGAAGTGAACACGGATTCGGAACCGGA 360

Qy 611 gttgtaaatgtaacctatttccagtaaggaccaaagctagacaagcactagacaaactgaat 670
|||||
Db 361 GTTGTAAATGTAACCTATTTCAGTAAGGACCAAGCTAGACAAGCCTTAGACAACTGAAT 420

Qy 671 ggaattcagttagaataattccaccttgaaagtagcctatccctgatgaaatgagccgac 730
|||||
Db 421 GGATTCAGTTAGAGAACTTCCCTTGAAGGTTGCCTACATCCCATGAAACTGCTGTC 480

D	b	121631	TAAAGAGAAATTTTGTAGTCTCTAAAGAGAGGTTAAACTTTGAAGCTTATATACAGT	121572
Q	y	1732	gccatcctttgctgctggcagagttattggaagaggagcacaacggtgaatgaacttca	1791
D	b	121571	GTCAATCCTTTGCTGCGGAGTTATTGGAANAAGGAGCAACCGTGAATAACTTCA	121512
Q	y	1792	gaattgtcgaagtcagaagtgttgccctcgtgaccagacacctgatgaaatgacca	1851
D	b	121511	GAATTTGTCAAGTCGCAAGTTGCTGTCCTCATGACACAGACACCTGATGAGAATGACCA	121452
Q	y	1852	agtggttgcaaaataactaggtcacttctatgcttgccagggtgcccagagaaaaattca	1911
D	b	121451	AGTGGTTGTCAAAATAACTGGTCACTTCTATGTCTGCCAGTTGCTTCAGAGAAAATTCA	121392
Q	y	1912	ggaaattcgtactcaggtcaagcagcaccaacaacagaaggtcctgcaagtggaccacc	1971
D	b	121391	GGCAATTCTGACTCAGGTAAACAGCA-CAACAGCAGAGGGTCTGCCAAGTGGACCACC	121333
Q	y	1972	tcagtcaagacggaagttaaagggtcagggaacgacgccacacagagggcagatgccaaacc	2031
D	b	121332	TCAGTCAAGATGGAAGTAAAGACTTCAGAAACAGGCCACCACAGAGCGAGATGGCAAA	121273
Q	y	2032	aaacacagattgcttaaccacacagatggcgctgacccccctaccagaatcacatgcaca	2091
D	b	121272	AAAGACAGATTGCTTTAACCAACAGACAGCGCGTGACCCCCATTCAGAGAATCATGTCACA	121213
Q	y	2092	agttttacctagccagttgttcttgaggaccaggcaacttttgaactcctgctctgtg	2151
D	b	121212	AGTTTTTACCTAGCCAGTTGTTCTGAGGACACAGGCACATTTTGAACCTCCTGTCCTGNG	121153
Q	y	2152	agaaatgatactttatgctctctgaaatgatatgacacccagcttttaaaacaaacaaaca	2211
D	b	121152	AGAAATGATATCTTTATGCTCTCCGAAATGATGACACCCAGCTTTAAAAACA-----	121100
Q	y	2212	acaaacaaaaaagggtgggggggaggggaagagaagagctctgcactcccttgtt	2271
D	b	121101	--AAACAAAAGAGGGTGGGGGAGGGAGGGAAGAGAGAGCTCTGCACCTCCCTTTGTT	121044
Q	y	2272	gtagctcacagataacagatatctaatcttcttaattcttcaattcccccataatgccagaa	2331
D	b	121043	GTAGCTCACAGTATAACAATATCTTAA---TCTTTAATATTTCCCCCAATAATTCAGAA	120987
Q	y	2332	attggcttaatgatgctttccactaaattcatcaaaagattgctctcctaatacctaattgtt	2391
D	b	120986	ATTGSCCTACTGATGCTTTTCACATAAT--TCAAGTAGATTGCTCCTAAATTCATTTGTT	120930
Q	y	2392	aaaattggatcagaataattatcacagggaacttaaatgttaagccaattagcatgaaaaa	2451
D	b	120929	AAAAATTGGATCAGAATAATTTATCACAGAAGCTTTAAATGTGTTAAGTATAGCATAGAAAA	120870
Q	y	2452	ctgtctcagtttatttttacctaaactaacatgaataaactaaggaaatgctgaat	2511
D	b	120869	CTGTTCTCAGCTTTATTTTACCTTACACGTGACATGAGTAACCTAAGGGAAGTGTGAAT	120810
Q	y	2512	gggtgtggcaggggtattaaacgtcatttttactcaactcactcaggtatttcagtaata	2571
D	b	120809	GGCTTGGCAGGTGATTAAACATGCATTTTATTCACTACCTCAGGTATTTCAGTAATA	120750
Q	y	2572	caatgaagaacaaattgttccctt-----	2596
D	b	120749	CAATGAAAGCAAAATTTGTCCCTATTGGCGAAATTCATCCCGAATATTTTCATGTAGGTT	120690
Q	y	2597	-----	2596
D	b	120689	GTTTTTCATTTTCCCTAAGTGTGAGTGTGCTGAGAAATAAAGGGACAGAGTACAAAAA	120630
Q	y	2597	-----	2596
D	b	120629	GAGAAATTTTAAAGCTAGGTGCCCGGGGGAGACATCACATGTCAGCAGGTTCTGTGATGC	120570
Q	y	2597	-----	2596
D	b	120569	CCCGAGCTGTAAACCAACAATTTTTTTTATTAGTGTATTTTCAAAAGGGGAGGAGTGTATG	120510

QY	2597	-----	2596
Db	120509	AAATAGGGTGTGGGTACAGAGATCACATGCTTCACAAGGTAATAAGATATACAAGGTAA	120450
QY	2597	-----	2596
Db	120449	ATGGAGGAGGGGAGATCACAGGACCACAGGACTGGGGCGAANAATPAAAATTCGTAAATGA	120390
QY	2597	-----	2596
Db	120389	AGTTTCAGGCAGGCATTTGTCATTGATACATCTTATCAGGAACACAGGTTTGAGAGCAGA	120330
QY	2597	-----	2596
Db	120329	CAACCGGTCCGACCACAAAATTTATTAGCGGGAGTTGCTCGTCTTAATAAGCCTGGGAAT	120270
QY	2597	-----	2596
Db	120269	GCTACGGGAGACTGGGGCTTATTTCATCCCTACAGCTACGACCATAAAGACAGCCGCT	120210
QY	2597	-----	2596
Db	120209	CCAAAGCCCTCAGGACACATTTTCTCTCAGGATGTTCTTGCTGAGAAAAAGAAAT	120150
QY	2597	-----	2596
Db	120149	CAGTGGTATTTCTCCCATTTCTTTTGAAGAGAGAAAATATGGCTCTGTTCCACCCAGCT	120090
QY	2597	-----	2596
Db	120089	CACCAGCAGTCAGAGTTTAAGGTTATCTCTCTGTTCCCTGAACATTGCTGTATTCTCTGT	120030
QY	2597	-----	2596
Db	120029	TCTTTTTTCAAGGTGCCAGATTTTCATATTGTTCAAACACACATGCTGTACAAACAATTT	119970
QY	2597	-----	2596
Db	119969	GTGCAGTTAACACAAATCATCATCACAGGTCCTCAGGTGACATACATCCTCCTCAGCTTATGA	119910
QY	2597	-----	2596
Db	119909	AGATGATGAGATTTAAGAGATTTAAAGACATAGGAATCACAAAGGTTATTGATTGGGAAGT	119850
QY	2597	-----	2596
Db	119849	GATAAGTCTCCATGAATCTTCAAAATTTATGTTTCAGAGATTGCAGTAAGACAGGCATA	119790
QY	2597	-----	2596
Db	119789	AGAAATATAAAAGTATTAAATTTTGGGAACATAATAATGTCCATGAAATCTTCACAATTT	119730
QY	2597	-----	2596
Db	119729	ATGTTCTTCTGCATGGTTTCAGCGGTCCCTCCATCNGGGTCCCTGCATTCGGCGAACA	119670
QY	2597	-----	2647
Db	119669	GTCTCTAAATTTTGAAAAATTTTATATACTTTATAATGATAGAAGTCCAATCATATTTT	119610
QY	2648	aaaaataaatttaaaatttaacagcaatcagctaaacaggcaaatgaagatttttactctct	2707
Db	119609	AAAAATAAAATTTAAATGTAAACAG-AAATCAGCTAACAGGCCAAATTAAGAAATTTTACTTCT	119551
QY	2708	ggcTggtgacagtaaaagctggaaaaattaaattcagggttttttgaggctttttgacacagct	2767
Db	119550	GGCTGATGACAGTAAGACTGGAAAAATTAATTTTTCAGGGTTTTTTCAGGCCTTTGACACAGT	119491
QY	2768	tattagttaaatcaaatgtttcaaaaaatacaggagcagtgacctagtaactctggagagcagac	2827
Db	119490	TATTA-RAAAAATCAAAATGTTCAAAAGATATGGAGCAGTGCCTAATATCTCGAGAGACGAA	119432

Qy	2828	tacatttattcttcoatttattatagttggaagatttttgaacggtactacaagaagtggctg	2887
Db	119431	TACCAATTATTTCTTTTCATTATATAGTTTGAACAAAGCTGTTGATGTGCTACTAACAAGTGGTGG	119372
Qy	2888	caggagattttgaaagcgctgggtttaaaggcttcaggagacttcagtttttggttttgac	2947
Db	119371	CAGGAGCCTTTGGACCGCTGGTTTAAATGGCTTTAGGAGACTTCAGTTTTTGTGTTTACG	119312
Qy	2948	tacatgattgaatgcataataaagctcttgtgtctcttgactatcaataacctataaagaagt	3007
Db	119311	TACATGATTGAATGCATAATAAATGCTGTGTGTTTTTGACCTTATACTAGAGAAAGT	119252
Qy	3008	gcataagtgaaagatgcagaacttcaactgactgctgcaaaagaagcttttagctttgct	3067
Db	119251	GCATPATTGGAAGAGATGACGAGCACTTTCAACTGAAATGGCAAAAAGCAAGCTTTAGCTTTGTC	119192
Qy	3068	ttatagatgcttagtttgcactcacacttcagaccaatggacagctcatagatgctg	3127
Db	119191	TTACAGGTTGCTTTAGTTTGGCAGTACACTTCAGACCCAATGGGACAGTCAATGATGCTGG	119132
Qy	3128	acagtgtttaaagcaaaaaagctcacatttccatggggccagcactgcatgagcttc	3187
Db	119131	ACAGTGTTTAAAGSCAACAAAAGGC-ACATCTCCATGTGCTGGCACTGTCATGAGCCTC	119073
Qy	3188	actaagctattttgaagatttttaagcactgataaattaaaaaiaaaaaaataatag	3247
Db	119072	ACTAAACATATTTTGAAGATTTTAAAGCAATGATAAATTTT-----AAAAAATTAG	119022
Qy	3248	actccacttaagtag---taagtatcaacaggatttctgtatactgtgcactcagctc	3303
Db	119021	ACTCTGCCTTAAGTAGTATATAAAGTATAGCAGGATTTCTGTATACTTCGCAATCAGTTTC	118962
Qy	3304	-tttgaaaaaaagtcacaaagatagagaatacagaagaagttttngggataaaattgaa	3362
Db	118961	TTTTTTTTAAAAAGTCAAAGGTTAGAGAAATACAAGAAAAAGTTTTTGGGATATAATTGGAA	118902
Qy	3363	tgactgtgaaaaacatatgaactttgatacgaactcattgtcactcctt-----gaca	3417
Db	118901	TGACTGTGAAAC-ATGACCTTTTGATCTAGAACTCATTTGTCTCACTCTCTTAACGGACA	118844
Qy	3418	gcaagccacagtcagtacaaattgctgggtgtgggtgtctccaaagccacgctgctct	3477
Db	118843	GCAAAGTCCAGTATGTACAAATTTGTTGGGTGTGGGTGGTGTCCAAGGCCACACTACTCT	118784
Qy	3478	ctgaattgatttttgagtttggnttgnaaagatgatcacagncatgttgcactgatott	3537
Db	118783	CTGAATTCAGCTTTTGAG--TTTTTTTGTGAAGATGATCAGTCATGTTACACTGACCTG	118727
Qy	3538	naaggacataintataaccctttaaaaaaaatccctgcctcattcttatttcgagaga	3597
Db	118726	AAGG-----ATATATATCCCTTTAAAAAATCACTCTGCTCATCTTATTTCAAGA	118675
Qy	3598	tgaatttcgatacacagactagatgtcttct-----gaagatcaattagacaattntgaa	3650
Db	118674	TGAATTTCTATACAGACTAGATATGTTTTTCTGAAGATCATATCAATTAGACATTTTGAA	118615
Qy	3651	aatgatttaaagtgttttcccttaagtctctcgaacaaagtttcttttgtagtttttaac	3710
Db	118614	AATGATTTAAAGTGTTCCTTAATGTTCTCAGAAAAACAAGTTCTTCCTTTGTAGTTTTAAT	118555
Qy	3711	caaaaaagtgcccttttggtcactggtttctctcagctatcatgatttttttttcacaca	3770
Db	118554	CAAAAAAGTGCCTTTTTTGTGCTACTGGATCACCTAGCATTCAT-AATTTTTTTTTTACATA	118496
Qy	3771	atgaattaaaaatgctaaaaatcatgagactggtttctggttgaggatttcaggtaagatgtg	3830
Db	118495	ATGAATTTAAATTAATCTAAAAATCATGACGCTGGCTTCTGTTGGATTTTCAGTGAGATGTG	118436
Qy	3831	tttaagccacagactttctccagtatattgatttttttccccaattatgatttttttaaaa	3890
Db	118435	TTTAAGGCCAGAGCTTTTCTCAATATTTGATTTTTTTTCCCCCAATATTTGATTTTTAAAAA	118376
Qy	3891	atatacacatagagagctgcatttaaaccctgctgggtttaaaattctgtcanatttcacttc	3950

Db	118375	ATATACACATAGTGTCTGCATTTT-ATATCTGCTGGTTTAAATCTATTATATATTTCACTTA	118317
Qy	3951	tagccttttagtaggcnaatcnaaattacttttacttaagcatttgtaattggagta	4010
Db	118316	TAGCCTTTTAGTATGGCAAAATGATTTTACTTTTACTTAAGCATTTGTAAATTTGGAGTA	118257
Qy	4011	tctggtactagctaagaaataattcnaataattgagtttctgactcncacaaanaatgggtcca	4070
Db	118256	TCTGCTACTAGCTAAGAAATAATCTATAAATGAGTTTGTACTCATCATATATGGATCA	118197
Qy	4071	ttctcatgnaataatgtgtcccccaatgcaggttctatttccaganaacctgaacacagat	4130
Db	118196	TTCCFCACTATAATGTGCCCAATGAGCTTTCATTTTCCAGATACCCTTCCACGCAGGAT	118137
Qy	4131	aaatttttcatcatttaggtcccccacaaaaaaacaaaaaaacaaaaaa	4181
Db	118136	AAAGTTTTTTCATCATTTAGGTGCCAAAAAAGAAAAAGAAAAAATATTCAA	118086

RESULT	8	
AC024199/c		
LOCUS		
DEFINITION	AC024199 188273 bp DNA HTG	12-AUG-2000
	Human sapiens chromosome 4 clone RP11-795O24, WORKING DRAFT	
	SEQUENCE, 24 unordered pieces.	
ACCESSION	AC024199	
VERSION	AC024199.3 GI:9798101	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Human sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 188273)	
AUTHORS	Waterston,R.H.	
TITLE	The sequence of Homo sapiens clone	
JOURNAL	unpublished	
REFERENCE	2 (bases 1 to 188273)	
AUTHORS	Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (25-FEB-2000) Genome Sequencing Center, Washington	
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
	MO 63108, USA	
COMMENT	On Aug 12, 2000 this sequence version replaced gi:9309539.	

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0795024
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 99% of reads
Chemistry: Dye-terminator Big Dye; 1% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 176662 bases at least Q40
Consensus quality: 179515 bases at least Q30
Consensus quality: 180915 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 185973; sum-of-contigs
Quality coverage: 5.02 in Q20 bases; agarose-fp
Quality coverage: 5.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1759; contig of 1759 bp in length

```


Query Match	26.7%;	Score 1117;	DB 5;	Length 3147;
Best Local Similarity	73.7%;	Prod. No. 4.1e-215;		
Matches 1622;	Conservative	0;	Mismatches 510;	Indels 69; Gaps

Qy	211	tttttctcgattttttttaaacacacactcttcacaaatgaacaaactgtatatcgg	270
Db	932	TCCTTTTAGCCTTTTATTGGGTTTTTTTTTTTATTATTACGATGAACAAAGCTGTAATATGG	991
Qy	271	aaacctcagcgagaaacgcgcgccttcgacctagaaaagtattctcaaggaacgcacaagt	330
Db	992	AAACCTAAGCGAAACGTCAGCCCCCGGATTTGGAAAGTCTCTCAAGGAGTCGAAGAT	1051
Qy	331	cccggtgtcgggaccttctcgtggaagactggtctacgcgttcgtggaactgcgcggacga	390
Db	1052	CCCTTTTCACGGGCAGTTCTTGCTCAAAAGTGGATACGCGCTTTGTGACTGTCCGGATGA	1111
Qy	391	gaactgggacctcaaggccatcgaggcgcttcacgtagtaaaatagaacactgcacgggaaacc	450
Db	1112	GAGCTGGGTATGAAGGCCATCGACACCTCTCAGGGAAAGTGGAGTTCGATGGCAAAAGT	1171
Qy	451	catagaagttgagcaactcggtcccaaaaaggcaaaagattcgaaaacttcagatacga	510
Db	1172	GATAGAAGTTGAACACTCAGTACCCAAAAGCAAGAGTCGAAGAGCTTCAGATAAGGAA	1231
Qy	511	tatccgcctcatattacagttggaggtgctgggtagtttactagtcagctatggagtggt	570
Db	1232	CATACCACCTCATCTACAGTGGGAGGTACTGGACACGCTATTAGCACAGTATGGGACAGT	1291
Qy	571	ggaagctgtgagcaagtgaacactgactcggaaactgcagttgcagttgtaaatgaacctattc	630
Db	1292	GGAAAACTGTGAGCAAGTTAAACACTGATTGAGAAACTGCGAGTAGTAGTAATGTAACATATGC	1351
Qy	631	cagtaaggaccaaactagacaagcactagacaaactgaatggatttcagttcagtagagaattt	690
Db	1352	CAATAAGGAGCAGCTAGACAAGGACTCGAAAAATTAACGGGTATTCAGCTTGAGAACTA	1411
Qy	691	caacttgaagttagctctatccctgatgaattggccgcgcagcaaaaaccccttgcagca	750
Db	1412	TAGCCTGAAAGTGACGTATATACCTGATGAATGGCAACACCGCAATCACCATCCCAAGCA	1471
Qy	751	g-----ccccgaggtgcgcggggcttgaggggcttgaggcagaggggctc	786
Db	1472	GCTTCAGCAGCCGACGACGCAACACCTCAAGGGAGGCGTGGGTTTGGGCAAAAGGGGACC	1531
Qy	787	ctcaaggcagggtctccaggatccgtatcccaaggcagaaacc---atgtgatttgccctct	843
Db	1532	TGCCAGGCAGGGGTCCCTGTGCTGCTGCAAGACCTTAAGCCCTCAGTCAGAGGTTCACCT	1591
Qy	844	gcgcctgcgtgtccaccaccaatttgttgagccatcattaggaagaaagaggtgccaccat	903
Db	1592	GAGGATGCTGTGCCCCACACAGTTTGTGGGGCGATCATTTGAAAAGAGGGTGCACCAT	1651
Qy	904	tcggaaacatcaccaaacagaccagctcaaaatcgatgtccacgcgtaaagaaatgcggg	963
Db	1652	CAGGAACATCACTAAACAGACCCAGTCAAAAATAGATATTCACCGTAAAGAAATGCTGG	1711
Qy	964	ggctgtgagaatcgattactatctctctactctgtaaggcagccctctgcggtctgttaa	1023
Db	1712	TGCTGCAGAGAAACCGATTACGATTCACTCAACCCCTGAAAGGCTGTTTCAGCTGCATCGAA	1771
Qy	1024	gtctattctggagattatcatagaagctcaagataataaaattcacagaagatccc	1083
Db	1772	GATCATTTGGAGATCATGCAGAAGGAAGCTCAGGATACCAGTTCTACTGAAAGAAATGCC	1831
Qy	1084	cttgaagattttagctcataaactttgttgacgctcttatttggttaaaagaaagaa	1143
Db	1832	CTTAAAAATATTAGCACATAACAATTTGTGTGCCGCTTATTTTGGGAAAGCAAGGAAGGAA	1891
Qy	1144	tcttaaaaaaattgagcaagacacagacactaaaaatcogatatctctcattgcagggaatt	1203
Db	1892	CCTTAAGAAATAGAGCAAGACAGATACCAAAATCACAAATCTCCACTACAGGACTT	1951

QY	1204	gacgctgataataatccagaacgcactattacagtgtaaaggccaattgtagacatatgtgccaa	1263
DB	1952	GACACTGTACAATCTCTGAACGGACCACTTACAGTAAAGAGCGCAGCATTTGAGACATTTGCGCAA	2011
QY	1264	aqctagaggagagatcatgaagaataatcaggagctcttatgaaatgatattgcttctat	1323
DB	2012	AGCCGAGGAAGAGTCATGAAGAAAAATTAGGGAGTCTCTATGAAGATGATATTGCTGCTAT	2071
QY	1324	gaatcttcaagcacatttaattcctggattaataatctgaacgccttggtgctgttccccac-	1382
DB	2072	GAATCTACAGGCACATTTGATTCTTGTTTAAACCTGAATGCATTTGGGCCCTTTTCCCAACC	2131
QY	1383	--ccacttcaggagatgccacctcccacctc- --agggcccccttcagccatgaactcctcc	1437
DB	2132	ATCTCTGTCAGGAATGCCACCACTCTGCTGAGGATTTCTCTCCGACAAACATCTGTGTTTC	2191
QY	1438	ctaccgcgagttg- --acgaatcagaagaacgagagactgtttcatcagtttatccagctctc	1494
DB	2192	TTATCCACCATTTTGGCAGCAGCCAGAGTCAGAGACTGTTCACTCTTCATCCAGCCAGCTTT	2253
QY	1495	atcagtcggtgcatcatcggcgaacgagggccagcacatcaaacgagcttctcgcgtttgc	1554
DB	2252	AGCTGTGGAGCANATTATTGGAAGCAGGACAAACACATCAAAACACTGTCTCGGTTTTTGC	2311
QY	1555	tggagcttcaattaaagattgctccagcgaagcaccagatgctaaagtggagatggtgat	1614
DB	2312	AGGAGCATCTATTAAAGATTGCACCCGCAGAGGGACCTGATGCCAAACTAAGAATGGTGAT	2371
QY	1615	tatcactgacacacagaggtcagttcgaagctcgggggaagaatttatggaaaaataa	1674
DB	2372	CATCACTGACGACCTCTCTGAAGCCCAATTATTAGGCTCAAGGAAGGATCTATTGGTAAACTTAA	2431
QY	1675	aqagaagaacttgtagtcttaagaagaagtggaacttggaactgcatacatcagatgccc	1734
DB	2432	AGAAGAAAACCTCTTTGGACCTTAAAGAAAGTGAACATTGAGGCCACCAATTAAAGTGCC	2491
QY	1735	atccttgtcgtcggcagagttatttgaaaggaggcgaacacggtgaactgcagaa	1794
DB	2492	GTCATATGCTGTGGACGTGTATTGGCAAAGGAGGCAAAACAGTAAATCAACTTCAGAA	2551
QY	1795	tbtgtcaadtcagaagttgtgttcctcgtgaccagacacacctgatgagaaatgacacagt	1854
DB	2552	TTTAAACAAGTGCAGAAGTTGTTGTGCGCGTGATCAACACCAGATGAAATGATCAAGT	2611
QY	1855	ggttgtcaaaaataactggtgcacttctatgcttcgcaggttgccagagaaaaattcagga	1914
DB	2612	GGTGTGTTAAGATAACGGGTCACTTCTATGCTAGCCAGCTTGCACAAAGGAAAATTCAGGA	2671
QY	1915	aattctgactcaggttaa- ----agcagcaccaacacagaagcctctgcaaaagtgcacc	1968
DB	2672	ATACTGTGCTCAGTTAAGAAGACAGCAGCAGCAACACAGCAAGACACGCGCAAGATGGACA	2731
QY	1969	acctcagtcacaacggaagttaaaggctcagg- aaacagcccacccacagagggcagatgcc	2027
DB	2732	ACCCCAACCAAGAAGAAAAATAAGGTTTCAGAAAAATGCGCCTTTCACAGACACAGATGCCA	2791
QY	2028	accacaaagcagatgcttctaaccacaagatggtgcgtgacccccctatccagaatcacatg	2087
DB	2792	GACCAAAGACACTGAGATATAAGACAGGTTCACACCTCTGCTGTGCTACGTGGGAGGAA	2851
QY	2088	cacaagttttacctagccagttgttcttgagaccaggaacattttgaaactcctgtctc	2147
DB	2852	AACACCTAGTACATCCCTGTT- ----TGAGACCACGGAACCACTTGAAC- ----TGTC	2901
QY	2148	tgtggaatgtatacctttatgctctctgaaatgatlgacacccagcttttaaacacaa	2207
DB	2902	TCTGAGAAATGATTATTAGGCTCTC- AAAACTTTTGAACCCAGCTTTAA- --AAATGG	2957
QY	2208	acaaacaaacaaaaaagggtgggggagggaggggaagagagagctctgcacttccctt	2267
DB	2958	AAAAAAAAGGCAAAAAAGAAATGAAGAAAAAATAAGGTCGCCCTTCACATTCGCTTT	3017

Qy	2268	tgttgtagctcacagataacagata--tctcaattcttcttcaatt--c 2315
Db	3018	TGATCTATTCACAAAGTTACATAAAACCTTTGTAGTCTTTTTTATTGTTCTTATCC 3077
Qy	2316	cccataatgccagaattgcctaataatgatctttcactaa 2356
Db	3078	CCCAACAATTTAGAAGCTTTCTTAAGAGCTTCTCTTA 3118
RESULT 10		
AF064633		
LOCUS	AF064633	2697 bp mRNA VRT 03-JUN-1998
DEFINITION	Xenopus laevis Vg1 RNA binding protein variant A mRNA, complete cds.	
ACCESSION	AF064633	
VERSION	AF064633.1	GI:3172446
KEYWORDS		
SOURCE	African clawed frog.	
ORGANISM	Xenopus laevis	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.	
AUTHORS	Havin,L., Glt,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Visraeli,J.K.	
TITLE	RNA binding protein conserved in both microtubule and microfilament-based RNA localization	
JOURNAL	Genes Dev. (1998) In press	
AUTHORS	Havin,L., Glt,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Visraeli,J.K.	
TITLE	Direct Submission	
JOURNAL	Submitted (11-MAY-1998) Biochemistry, Cambridge University, Tennis Court Road, Cambridge CB2 1GA, UK	
FEATURES	Location/Qualifiers	
source	1..2697	
	/organism="Xenopus laevis"	
	/db_xref="taxon:8355"	
	/dev_stage="oocyte"	
	299..2080	
CDS	/note="Vg1-RBP; Vg1 localization element binding protein"	
	/codon_start=1	
	/product="Vg1 RNA binding protein variant A"	
	/protein_id="AAC18597.1"	
	/db_xref="GI:3172447"	
	/translation="MNKLXIGNLSENVSPDLSLFKESKIPFTGQFLVKSGYAFVDC PDETFWAKAIDTSLGKVELHGKLVIEVHSVPKQRKQLQIRNIPHLOWEVLDSLLA QYGTVCNEQVNTDSAVVNVYANKHARQLEKNGQLENLSUKVTYIIPDEHAT PQSPSOLOQPOHSPGFGORGPARGQSPGAARPKQSEVFLRMLVPTQFQGA IIOKEGATRNITKQTSKIDIRKENAGAAEKPIITHSTPEGCSAAKTIIMEIQKE AODKTEETPELKLAKHNFGVLKIGKGNLKIIEQDVTDKITISPLQDLTLYNPER TITVKSIEFCAAKEEVKKK IRESYENIETAAANLQHLIPGLNALGIFPPSSSGM PPSGAVSPTTSASVPPPGQPESETVHLIFALAVGAILIGQGHIKQLSRFPAGS IKTAEPGDAKLUMIYITGPPAQKAGRIYIKLKEENFFGPRKEEVLEAHIKVPS YAAAGYIRGGKTVTNLQNLSAEVVVPRDTPENDQVVVYKITGHFYASQLAQRKIQ EILAAVRRQQOQKQAQSQGPQPRK"	
BASE COUNT	855 a 572 c 588 g 681 t	1 others
ORIGIN		
Query Match	26.7%	Score 1116.4; DB 5; Length 2697;
Best Local Similarity	73.2%	Pred. No. 5.4e-215;
Matches 1602;	Conservative 0;	Mismatches 531; Indels 57; Gaps 11;
Qy	211	tttttttcgttatatttttttaaacaccactcttcacaatgaacaaactgtatcgg 270
Db	259	TCTTTTAGCCCTTTATTGGTTTTTTTTTTTATTTTACGATGAACAGCTGATATTGG 318
Qy	271	aaacctcagcagaacccgcctcggaacctgagaagtattcttcagacgcgaagat 330
Db	319	AAACCTAAGCGAAACATGCTAGCCCCCGAATTGGAAAGTCTCTTCAAGGAGTCGAAGAT 378
Qy	331	cccgggtcgggacccttctctggtgaagactcgcctacacqcttcgtgacqtcgcccggaqca 390

Db	1459	ATCCTCGTCAGGAGTCCACACACATCTGCTGGAGTGTTCCTCTCCGACACACATCTGCTTC	1511
QY	1438	ctaccgcgagttgt---agcaatcagaaacgagactgttcatcagtttatcccgactct	1494
Db	1519	TTATCCACCATTTGGGCAGCAGCCAGATGTCAGAGACTGTTCATCTCTCATCCGAGTTT	1578
QY	1495	atcagtcygtgcatactatcggcgaacagggccagcacaatcaagcagctttctcgtcttgc	1554
Db	1579	AGCTGTGGAGCAATTTATTGGAAGCAAGAGCAACACATCAAAACAACGTGCGGCTTTGC	1638
QY	1555	tgagcttcaattaagattgctccagcgggaagcaccagatgctaaagtgagatggtgat	1614
Db	1639	AGGAGCATCTATTAGATTGTCACCCGAGAGGACCTGATGCCAAACTAGATGGTGAT	1698
QY	1615	tatcactggaccaccagagctcagttccaaagctcaggaagaatttatggaataataa	1674
Db	1699	CATCACTGGACCTCCTGAAGCCCAATTTAAGSCTCAAGGAAGGATCTATGGTAAACTTAA	1758
QY	1675	agaagaaaactttgttagtcctaaagaaggtgaaacttgaaactcatcatcagatgccc	1734
Db	1759	AGAAGAAAACTTCTTTGGACCTTAAAGAGAAGTAGTGAACCTTGAGGCCACATTAAGTGC	1818
QY	1735	atcctttgctgctgcagagattatggaagaggagcaaaagtgatgaacttcagaa	1794
Db	1819	GTCAATATGCTGTGGACGTGTATTGGCAAGAGGCAACACAGTAATACTGAACCTTCAGAA	1878
QY	1795	tttgtcaagtgcagaagttgttcctcgtgaccagacacctgatgagaatgaccaagt	1854
Db	1879	TTTAAACAGTGCAGAGTTGTTGTGCCCGTGATCAACACACCATGAAATGATCAAGT	1938
QY	1855	ggttgcataataactggctactcttatgcttgcaggttgcgcagagagaaaaattcagga	1914
Db	1939	GGTTGTTAAGATAACGGGTCACTTCTATGCTAGCCAGCTTGCACAAAGGAAAAATTCAGGA	1998
QY	1915	aatttgactcaggtaa-----agcagcacaacaacagaaagctctgcagagtggacc	1968
Db	1999	AATCACTGGCTCAGGTTAAGAAGACACGACGACACAGCAGAAAGACACGCAAAAGTGGACA	2058
QY	1969	acctcagtcaagcaggaagttaaaggtctcagg-aaacagccccaccacagagcagatgcc	2027
Db	2059	ACCCCAACCAAGNAGANAATAAAGGTTCAGGNAATGGCCCTTCACAGACAGATGCCA	2118
QY	2028	aaccaagacagattgctttaaccaacagatggcgctgacccccctatccagaatcacatg	2087
Db	2119	GACCAAGACAGACTGAGATAAAGACAGAGTTCACCTCTGTCTGTGCTGGGAGGAA	2178
QY	2088	cacaagttttaccttagccagttgttctgagaccagcagaacttttgaactcctgtctc	2147
Db	2179	AACCACTAGTATCCCTCTGTT----TGAGGACCGACGCAACCATTTGAAC-----TGTC	2228
QY	2148	tgtgagaatgtataactttatgctctctgaaatgtatgacccccagcttttaaacacaa	2207
Db	2229	TCGAGANAATGATTATTAGGCTCTC-AAAACCTTTTGAACCCAGCTTTAA--AAATGG	2284
QY	2208	acaaacaaacaaaaaggggtggggagggggaagagaagagctctgcacttccctt	2267
Db	2285	AAAAAAAGGCAAAAAAGAAAAATGAAAAAATAAGGTCCCTTCACCTTCCTCCCTTT	2344
QY	2268	tgtgtagctcacagataaacagataattcttaattcttcttaattcccccaataatgccc	2327
Db	2345	TGATCTATTCTCAAGATTAAACATAAACCTGTGTAGTCTTTTTTTTATTGTCTTATTC	2404
QY	2328	agaaattggcttaagtatgctttcactaaa	2357
Db	2405	CCCAACAATTTAGACTTTGTTTAAATGAA	2434
RESULT	11		
LOCUS	AP002376		
DEFINITION	AP002376	161404 bp DNA	PRI 21-JUN-2001
ACCESSION	AP002376	Homo sapiens genomic DNA, chromosome 11q, clone: RP11-349I16, complete sequence.	

VERSION	AB002376.3	GI:14517592
KEYWORDS	HTG.	
SOURCE	Homo sapiens DNA, clone:RP11-349I16.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (sites) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE	Homo sapiens genomic DNA	
REFERENCE	Published Only in DataBase (2000) In press	
AUTHORS	2 (bases 1 to 161404) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)	
COMMENT	On Jun 20, 2001 this sequence version replaced gi:11094148.	
FEATURES	Location/Qualifiers	
source	1..161404 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="11" /map="11q" /clone="RP11-349I16"	
BASE COUNT	47396 a 33453 c 33200 g 47355 t	
ORIGIN		
Query Match	25.9%;	Score 1084.4; DB 9; Length 161404;
Best Local Similarity	88.4%;	Reqd No. 1.4e-208;
Matches 1305; Conservative	0; Mismatches 127; Indels 44; Gaps 11;	
Qy 2730	aaattaattcagggtgttttggagcgttttgacacagttattagtt---aaatcaaatgt	2786
Db 130573	AAATTAATTTTCAGGGTTTTTTTTCAGGCTTTTCACACAGTTATTAGCTAAAAATCAAAATGT	130632
Qy 2787	tcaaaaatacaggacagtcctagatctctgagagcagcaactaccatttattcttcatt	2846
Db 130633	TCAAAGATATGGACAGTCGCCATAATGTCGTGGAGACAGCACATACCATTTATATTTCATT	130692
Qy 2847	tatagttgggaagttttttgacggtactaa-----caaaagtggtcgacaggagatttggga	2901
Db 130693	TATAGTTGGGAAAGTTTTTGATGTTACTACAGATCAAAGTGGTGGCAGGAGTTTTCGA	130752
Qy 2902	acggtggttttaaatggcttcaggagacattcagttttttgttagctacatgattgaatg	2961
Db 130753	ACGGCTGGTTTTAAATGGCTTCAGGAGACTTCAGTGTGTTTTGTTAGTCACATGATTGAATG	130812
Qy 2962	cataataaactgtttgtgcttctgactatcaatacctaaagaagaagtcatcagtgaaagag	3021
Db 130813	CGTAATAAATGCTTTGTGCTTTTGACTATCAATGCCTAGAGAAAAGTCATCAGTGAAGAG	130872
Qy 3022	atgcaagacttcaactgactggcgaagaagcaagcttttagcttctgtcttataggatgctta	3081
Db 130873	ATGCAGACGTTCACTGATTTGGCCAAAAGCAAGCTTTAGCTTGTCTTACAGGATGCTTA	130932
Qy 3082	gtttgccactacacttcagaccaatgggacagtcatagatggtgtgacagttgtttaacyg	3141
Db 130933	GTTTTGCCAGTACACTTCAGACCAATGGATAGTCATAGATGGCTGCAGTGTTTAAAGG	130992
Qy 3142	caacaaaggctacatttccatggggcagcactgtcatgagcctcactaagctatttgg	3201
Db 130993	CAACAAAAGGCTACATCTCCATGTGGCCAGACACTGTCATGAGCCTTCATAAGCTATTGTG	131052
Qy 3202	aagatttttaagcactgataaattcaaaaaaataaataaattagactccaccttaagt	3261
Db 131053	AAGATTTTAAAGCAATGATAAATTA-----AAAAAATTTATATACCCCACTCAAGT	131103
Qy 3262	aq----taaaqtaLaacaggaatttctgtatactgtgcaatcagttctcttggaaaaaagt	3317

Db 131104 AGTACATAAAGTAGACAGGATTTCCGTATATCTGCAATCAGTCTCTTGAAGAAAGT 131163
Qy 3318 caaaagatagagatacaagaagaatttngggataaatttgaatgactgtgaaacat 3377
Db 131164 CAAAAGTAGAGATAACAAGAAA--TTTTTGGGATATAATTTGAATGACTGTGAAAC-- 131219
Qy 3378 atgacctttgataacgaactcatttgcactc-----cttgacagcaagccagtaag 3432
Db 131220 ATGACCTTTGATAGTAACTCACTTTGTGCTACTCTTAACTTGACAGCAAGCCGAGTATG 131279
Qy 3433 tacaattgtgtgggtgtgggtgtctcccaaggccagctgctctcgaattgatttttt 3492
Db 131280 TACAACGTGTGTGACTGTGGTGTCTCCAAAGCCAGCTGCTCTCGAATGTATTTGA 131339
Qy 3493 gaggtttggnttgaagatgatcacagncatgtttacactgatctttnaaggacatatnta 3552
Db 131340 G---TTTTTGTGTGAAGATGATCACAGTCATGTTTACACTGACCTGAAGGCATATATATA 131396
Qy 3553 taaccttttaaaaaaaatccccgcctcattcttatttcgagatgaattcgatacag 3612
Db 131397 CATAACACTTTTAAAAAAACACTCTGCCTCATCTTTATTTCAAGATGAATTTCTATACGG 131456
Qy 3613 actagatgtcttct-----gaagatcaattacacattntgaattgaattaaagt 3663
Db 131457 ATTAGATATGTTTTTCTGAAGATCATATCAATTAGACATTTTGAATAAGACTTAAAT 131516
Qy 3664 gtttccctaattgtctctgaaacaaagtcttctttttagtttttaacccaaaaaagtcccc 3723
Db 131517 GTTTTCTTAATGTTCTCAGAAAAAAGTTTCTTTTGTAGTTTAAACCAAAAAGTGCC 131576
Qy 3724 ttttgcactggtttctctctagcattcatgatttttttttttttttttttttttttttttt 3782
Db 131577 TTTTGTGCTAGTGGATTCACCTAGCATTCATAAATTTTTTTTTTTCATACAAATGAATTTAAAT 131636
Qy 3783 tgcataaatcatgactgagcttctctgtgtgattcaggttaaatgttttaagggcaga 3842
Db 131637 TGCTAAATCATGAGCTGGCTTCTGTGTGATTTCAAGGTGAGTGTGTTTAAGGCCAGA 131696
Qy 3843 gcttttcagttattgattttttttccccaatttttgcataatttttttttttttttttttttt 3902
Db 131697 GCTTTTCTCAATATTTGATTTTTTTTCTCCTCAATATTTGATTTTTTAAAAATATACATAG 131756
Qy 3903 gactgcatttaaaacccgtggttttaattctctgctaanattcactctcgtccttttagt 3962
Db 131757 GTGCTGCTATTT-ATATCTGCTGTTTAAATCTGTCTATATTTTCACTTCTAGACTTTTAGT 131815
Qy 3963 atggcnaatcanaatttacttttacttaagcatttgaatttgaattgattctgtactagc 4022
Db 131816 ATGGCAATCATATTTTACTTTTACTTTAAGCATTTGTAGTTTGGAGTATCTGGTACTAGC 131875
Qy 4023 taagaaataatcctnataattgattttgttactcnccaaanatgggtcattcctcatgnat 4082
Db 131876 TAAGAAATAATTTATAGTTTGTAGTTTCTGCTCACCATAATATGGATCATTTCTCATCTAT 131935
Qy 4083 aatgtncccccatgcagctctcattttccaganaaccttgacgagcataaatttttcat 4142
Db 131936 AATGTGCCCCAAATGACAGTTCATTTTTTCCAGATACCTTTGACGAGAGTGAAGTTTTCAT 131995
Qy 4143 catttaggtccccaaaaaataaaaaaataaaaaa 4178
Db 131996 CTTTGTAGTGCACAAAAAATAAAAAAAGGGCAAGA 132031

RESULT 12
AF064634
LOCUS AF064634 2405 bp mRNA VRT 03-JUN-1998
DEFINITION Xenopus laevis vgl RNA binding protein variant D mRNA, complete cds.
ACCESSION AF064634
VERSION AF064634.1 GI:3172448
KEYWORDS African clawed frog.
SOURCE

ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 2405)
AUTHORS Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE RNA binding protein conserved in both microtubule and microfilament-based RNA localization
JOURNAL Genes Dev. (1998) In press
REFERENCE 2 (bases 1 to 2405)
AUTHORS Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Pressman Schwartz,S., Standart,N.M. and Yisraeli,J.K.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Biochemistry, Cambridge University, Tennis Court Road, Cambridge CB2 1GA, UK
FEATURES
source
1..2405
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="oocyte"
379..2163
/notes="Vgl-RBP; Vgl localization element binding protein"
/codon_start=1
/product="Vgl RNA binding protein variant D"
/protein_id="AAC18598.1"
/db_xref="GI:3172449"
/translation="MNKLYIGNSENVSPTDLSLFKESKIPFTGQFLVKSQYAFVDC PDETWAMKAIDTSLGVELHGVIEVSHSVPKRQRKQIIRNPFIHLOWEVLDSLLA OYGTVCNQVNTESAVVNTVYANKEHAROGLEKINGVOLENYSIKVTYIDENAT PQAPSQLOQPOQHQPOQRGFGQPARQSGPAAARPKPQTEVPLRMLVPTQFVG AIIKGKATIRNTKQSKIDHRKENAGAEKPIITHSPEGCASAKILMEIMQK EAQDTFTTEIPKILAHNNFVGLJEGKNLKIETQDTHKTTISPLQDLFLYNPE RTTIVAGSEPCAKABEEMKIKRESYENIDAAANLQAHILPGLINLALGFLPSSSG MPSPVSGVPSPTSTSYPPFQOPESETVHLFIPALAVGAIIGQOQHIKQLSRFAGA STIKIAEPGDAKLRMVLTGPPEAFKAGRIYGLKKEFNFGPKEEVKLETHIKVP SYAAGRVIGKGTVNELNLTSAEVVVPDQTPDENDEVVKITGHFTYASQLAQRKI QEILAVRRQQQQQKTVSQGQPQRRK"
BASE COUNT 763 a 537 c 537 g 568 t
ORIGIN
Query Match 25.9%; Score 1084.2; DB 5; Length 2405;
Best Local Similarity 73.2%; Pred. No. 1.7e-208;
Matches 1521; Conservative 0; Mismatches 503; Indels 55; Gaps 8;
Qy 192 ggtgggtttaaatttcgttttctcgttatttttttaaaacaccactcttcacaa 251
Db 320 GGCTCGCTCTCTTTTAGCTTTTATTATTATTTTATTTTATTTTATTTTACAA 379
Qy 252 tgaacaaactgtatctcgaaacctcagcagaaacgcgcgcctcgcgacctagaaagta 311
Db 380 TGAACAAGCTGTATTTGGAACCTCAGCGAAAAACGTGAGCCCCACGATTTGGAAGTC 439
Qy 312 tcttaaggagcgcgaagatccccgggtgtcgggacccttccctgggtgaagacgtgctacgcgt 371
Db 440 TCTCAAGGAGTCTAAGATCCCTTTCACGGCGAGTTTCTTGGTCAAAAGTGATACCGGT 499
Qy 372 tctggaactcccgagcagagctggccctcaagccatcagggcgtcttccaggtaaaa 431
Db 500 TTGTGGACTGTCCGGATGAGACGTGGGCTATGAAGGCCATCGACACCTCTCAGGGGAAG 559
Qy 432 tagaactgcacgggaaacccatagaagttagcactcgggtcccaaaaggcaaggattc 491
Db 560 TGGAGTTGCATGGCAAAAGTGATTGAAGTCGAACACTCGGTACCCAAAAAGGCAAGGAGTC 619
Qy 492 ggaacttcagatacagaaatattccgcctcatttacaagtgagggtgctgtagatttac 551
Db 620 GAAAGCTTCAGATACGAACATACCCGCTCATCTACAGTGGGAGGTACTGCACAGCCTGT 679
Qy 552 tagtccagttatggagtgtgtgagagctgtgcaagtgaacactgactcgggaaactcgag 611
Db 680 TGGCACAGTATGGACAGTGGAAAACTGTGNGCAAGTGAACACTGAATCAGAACTGCAG 739

QY	612	ttgtaaatga	taaccta	tattccag	taaggccaagct	atagacaaga	ccactagacacaa	ctaagt	671
Db	740	TAGTGAAT	GTAA	CATATG	CCAA	TAA	TAGGAGCAGCGT	TAGCAAAAGGACTCG	799
QY	672	gatttcag	ttagaga	atttcac	cttgaaag	tagctata	tccctctgat	gaatggccqccc	731
Db	800	GCTATCAG	TGCGGA	AAACAT	ATAGCCT	GAAGTGCAC	TATATACCT	TATGTAATGGCAACAC	859
QY	732	agcaaa	ccccct	-----	-----	-----	-----	-----	764
Db	860	CGCAAGC	ACATCC	CAACAG	CTTCAG	AGCAGCGC	GACGACAC	ACCCCTCAAGGGAGGC	919
QY	765	gggggct	tgagcag	agggtc	ctcaag	gcagggt	ctccagat	cogtattccaagcaga	824
Db	920	GTGGGTT	TGGGCA	AGGGAC	CTGCCAG	GCGAGGAT	CCCTTGGTGC	TGTGCAAGACCTA	979
QY	825	aacc	---atgt	gat	tggcct	gtgcct	gtggttcc	ccaccacattt	881
Db	980	AGCCTCAG	ACAGAG	GTTCGCGT	GAGAAT	GCTGGT	TCCACACAC	ACTTTGTAGGGCGCATCA	1039
QY	882	taggaaaa	gaaggt	gcccacc	attcgga	acatcc	acaaacag	accagctctaaatogatg	941
Db	1040	TTGGGAA	ACAGGG	TGCTACCA	TACGAA	GCATCAGT	TAACAGAC	CCCGCTCAAAAATAGATA	1099
QY	942	tcaccg	taaa	gaa	aaatcg	ggggcgt	ctgaga	agtgcattactatcctctctactcctg	1001
Db	1100	TTCCCGT	TAAGAA	ATG	CAGGTG	TCGAGAA	ACCGATTAC	AAATTCATCAACCCCTG	1159
QY	1002	aaggcac	ctctcg	gctgt	taagt	tattct	tgagatt	tatgcataa	1061
Db	1160	AAGGCTGC	TCACTG	CATGT	TAAGAT	CATTAT	TGGAGAT	CATCSCAAGGAGGCTCAGGATA	1219
QY	1062	taaaa	ttc	acaga	agagat	ccccct	tgaa	gtatttagctc	1121
Db	1220	CCAAGT	TCTACT	GAAGAA	ATCCCT	TTAAAA	ATATTAG	CACATTAACAAAGCTTTGTTGGCGCTC	1279
QY	1122	tattg	taaa	gaag	gaaga	aatctt	aaaaaa	attgac	1181
Db	1280	TTATTGG	TAAAG	AGGAGG	AACCTTT	AGAAAT	TGAC	AGATACAGATACAAAAATCA	1339
QY	1182	cgatat	ctccat	tcag	gaattg	acgctgt	gataat	ccagaacg	1241
Db	1340	CAATAT	TCTCCACT	TACAGG	ATTGAC	ACTGTAC	ATATCC	AGACGACCATTTACAGTAAAG	1399
QY	1242	gcaat	gttag	acat	gtgcc	aaagct	tgagg	ggagatcat	1301
Db	1400	GCAGCAT	TGAGCC	ATGTG	CCAAAG	CCGAGG	AAAGAAAT	CATGAAAAAAATTTAGGGAGTCT	1459
QY	1302	atgaa	aatgat	attg	cttct	atga	attctc	aa	1361
Db	1460	ATGAA	AAATGAT	ATTGCT	GATGA	ATCTC	AGGCAC	ATTTGATTCTCTGGTTTGAACCTGA	1519
QY	1362	acgc	cttgg	ggt	ctgtt	ccccc	ccacttc	-----	1415
Db	1520	ATGCAT	TGGCGCT	TTTTCC	CATCAT	CTCAT	CAGGAAT	TGCCAC	1579
QY	1416	ccctt	acgc	atg	act	gact	ctcct	ccacg	1472
Db	1580	CTTCACT	ACTTCA	TCTACT	TCTTAT	TCCACCA	TTTGG	GCACGACCCAGATCAGAGACCG	1639
QY	1473	ttcat	ag	ttt	at	ccag	ctctat	cag	1532
Db	1640	TTCA	TCTT	CATCC	AGCTTT	AGCT	GTGG	AGGCGATTATTTGGAAACGAAGGACACACA	1699
QY	1533	tcaag	cag	attt	ct	gctt	gtg	agctt	1592
Db	1700	TCAAA	CAACT	AT	CAGT	TTTTGC	AGGAG	CATCTATTAA	1759
QY	1593	atg	ct	aa	ag	tg	ggat	tg	1652
Db	1760	ATGCC	AAACT	TAAG	ATG	TGAT	CTAT	CTAG	1819

Qy	1653	gaagaatttatgaaaaataaaagagaagaacatttggtagttcctaataaagagagggtgaaac	1712
Db	1820	GAAGGATCTTAACTTAACTGAAAGAAGAAAACCTTTCCGACACTTAAAGAAGAGTGAAAC	1879
Qy	1713	ttaagactcatatcagagtgcctcttggctggcgagtgattatgtgaaaaggaggca	1772
Db	1880	TTGAGACCACATTAAAGTCGCCGTATATGCTGTGGACGGGTTATTGGCAAGAGGAGAA	1939
Qy	1773	aaacggtgaaatgaacttcagaatttgtcaagtcagagaattgttgcctcgtgcaccaga	1832
Db	1940	AAACAGTAATGAATTCAGATAATTAACAAGTGCAGAGGTTGTTGTGCCCGTGATCAA	1999
Qy	1833	cacctgatgagaatgaccaagtggttgcaaaaaaatcgtgccaattctatgcttgcacgg	1892
Db	2000	CACCAGATGAAAAATGATGAAGTGGTCGTTAAGATAACGGGTCACTTCTATGCAAGTCAGC	2059
Qy	1893	ttgccagagaaaaattcaggaaattctgactcaggttaa-----agcagcaccaacaac	1946
Db	2060	TTGCACAAAGGAAATTCAGGAAATATTGGCTCAGGTAAAGAGACAGCAGCACACACAGC	2119
Qy	1947	agaaggtctctgcaagtgagccacacctcagtcagcagcgaagtaaagggtccagg-aaacag	2005
Db	2120	AGAAACACAGTGCAAAGTGCACAACCCCAACCCGAGAAGAAAATAAAGGTTTCAGGAAATGG	2179
Qy	2006	cccacacagaggcagatgccccaaacacacagacagattgcttaacacacagatggcgctg	2065
Db	2180	CCCTTCACAGACAGACATGCCAGACCAAGACACAGCATGAGATGAAGACAAAATTCACACTC	2239
Qy	2066	accctctaccagaatcacatcacagaagtttttacctagccagttgittctgagagaccag	2125
Db	2240	TTGTCCTCTGCTATGTGG----GAAGAATACCACTAGCGCATCCCTGTTTGAGGACCAG	2295
Qy	2126	gcaacttttgaactcgtctctgtgagaatgtatacttctgtctctctgaaatgtatga	2185
Db	2296	GCAACCAATTAAC TG---TCTCTGAGAATGATTAATTAGGCTTTCAAAACCTTTTGAA-.2350	
Qy	2186	caccagctttaaaaaacaaacaaacaaacaaacaaaaa 2224	
Db	2351	ACCACGCTTTAAAAATGGAAAAAAGGCCAAAAAAGAAA 2389	
<hr/>			
RESULT	13		
AF055923			
LOCUS	AF055923	2161 bp mRNA	VRT 18-MAY-1998
DEFINITION	xenopus laevis RNA binding protein Vera mRNA,	complete cds.	
ACCESSION	AF055923		
VERSION	AF055923.1	GI:3136157	
KEYWORDS	African clawed frog.		
SOURCE	Xenopus laevis		
ORGANISM	Xenopus laevis		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		
	Xenopodinae; Xenopus.		
REFERENCE	1 (bases 1 to 2161)		
AUTHORS	Deshler,J.O., Hightett,M.I., Abramson,T. and Schnapp,B.J.		
TITLE	A highly conserved RNA-binding protein for cytoplasmic mRNA		
JOURNAL	Localization in vertebrates		
MEDLINE	Curr. Biol. 8 (9), 489-496 (1998)		
AUTHORS	98228351		
TITLE	2 (bases 1 to 2161)		
JOURNAL	Deshler,J.O., Hightett,M.I., Abramson,T. and Schnapp,B.J.		
FEATURES	Submitted (26-MAR-1998) Cell Biology, Harvard Medical School, 240		
source	Longwood Ave., Boston, MA 02115, USA		
	Location/Qualifiers		
	1..2161		
	/organism="Xenopus laevis"		
	/db_xref="taxon:8355"		
	135..1919		
	/codon_start=1		
CDS	/product="RNA binding protein Vera"		
	/protein_id="AAC41285.1"		
	/db_xref="GI:3136158"		


```
QY 2078 gaatcacatgcacagtttttaccttagccagttgtttctgaggagaccaggaactttttaa 2137
    |||
Db 2008 TGTGG---GAAGAAACCACCATCGGCATCCCTGTTTGTAGGACCGACGCAACCATTTAA 2063
    |||
QY 2138 ctctgtctctgtgagaatgtatacttattctgtctctgtaaatgtatgacacccacttta 2197
    |||
Db 2064 CTG-----TCTCTGAGAAATGTTATTTAGGCTTTCAAACCTTTTGAAACCCAGCTTTAA 2118
    |||
QY 2198 aaacaaacaaacaaacaaacaa 2224
    |||
Db 2119 AAATGGAAAAAGGCAAAAAAGAAA 2145
    |||

RESULT 14
LOCUS AF161270 2011 bp mRNA VRT 27-JUL-1999
DEFINITION Danio rerio Vg1 RNA binding protein mRNA, complete cds.
ACCESSION AF161270
VERSION AF161270.1 GI:5596631
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 2011)
AUTHORS Zhang,Q., Yaniv,K., Oberman,F., Wolke,U., Cit,A., Fromer,M.,
Taylor,W., Meyer,D., Standart,N., Raz,E. and Yisraeli,J.K.
TITLE Vg1 RBP Intracellular distribution and evolutionarily conserved
expression suggest multiple roles during development
JOURNAL Mech. Dev. (1999) In press
AUTHORS Oberman,F., Fromer,M. and Yisraeli,J.K.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1999) Anatomy and Cell Biology, Hebrew University
Hadasah Medical School, POB 12272, Jerusalem 91120, Israel
LOCATION/Qualifiers
FEATURES
source
1. .2011
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="linkage group 19"
/map="between Z3023 and Z15451"
1. .1749
/note="Vg1 RBP"
/codon_start=1
/product="Vg1 RNA binding protein"
/protein_id="AAD45610.1"
/db_xref="GI:5596632"
/translation="MNKLYIGNVSEQASALDESIFEQWKIPFSLKVGAFVDC
PDEKVMRAIDTLGKVELHGKLVLEHVSVPKRSCKLOIRNPPHMQWELDGLLA
QYGTVESCEQNTDETAVNVVRGAKDQAREMDKLNGLFMYNALKVSIPTETAA
ADAPAVGRRGNFPRPQPSLGARPKQSDVPURLLVPTQFVGAILKEGATIR
NITQTSKIDIHRRKENAGAEKPTVHSTPEGSSACRNIMEIQEADIKTEEI
PLKILAHNNFVRLIGREKRLKIEQDTDKITISPLQDLTLNPERTIVKGTLD
CAKAEETMKVRESYENDVAAMHNLNPLGLNLALPFGAASGGISPSVYSGPP
PGAQGVQSGAQMSESVHLFIPALVANGALIGQGHILKLSRFAGASIKIADGI
DAKRMVISPPEAQKQAQRIFGKLEENFPKPEVKLEAHIKVPSFAAGRVIGK
GKTVNELQNTLSAEVVPVPRDQTFDENQVVVVKITGHFYASQLAQRKIQETISQVRQ
QQRPASAGPVPARRK"
BASE COUNT 581 a 494 c 501 g 435 t
ORIGIN
```

Query Match 20.2%; Score 845.8; DB 5; Length 2011;
Best Local Similarity 69.2%; Pred. No. 2.1e-160;
Matches 1220; Conservative 0; Mismatches 522; Indels 21; Gaps 4;

```
QY 251 atgaacaaactgtatatcggaacactcgagagaacgcgcgcgcctcggaacctagaaagt 310
    |||||
Db 1 ATGAATAAGCTATACATCGGGAAGTACGACGAGCGAGCGCGCTGACTTGGAAAGT 60
    |||||
QY 311 atcttcaaggacccaagatcccgtgtcgggacccttccttggtgaagactggtcacgcg 370
```


Db 1135 GCGAGTGGAGGATTTACACCTCTGTGGTATCAGGACCACCGCCTGGGGCTCAAGCTGGA 1194

Qy 1436 cctaccocgcagtttgagcaatcagaacagagactgttcatcagttattccccagctcta 1495

Db 1195 TATCAATCATTTGGGGCGCATGAGTGGGAAACTGTGCACCTTCTTATTCCTGCACTG 1254

Qy 1496 tcaatcggtgccatcatcggaacagggccagcacatcaagaacagctttctcgcttgcct 1555

Db 1255 GCAGTTGGTGTATTATCGGCAAGCAGGGCCCAACACATTAAACAGCTCAGCGCTTCGCT 1314

Qy 1556 ggaactcaatgaattgtctccagcggaagcaccacagctcgtctaaagtgaagatggtgatt 1615

Db 1315 GGTCCGTCAATCAAAATGGCCCTGCTGATGGAAATTGATGCCAAGCAAGATGGTTATT 1374

Qy 1616 atcactggaccacagaggtcgtcagttcaaggtcagggagaagaatttatgaaaaattaaa 1675

Db 1375 ATTTCTGGACCACAGAGGCACAGTTTAAAGGCTCAGGAGCGTATCTTTGGAAAGCTTAAA 1434

Qy 1676 gaagaaactttgttagtctaaagaagagtgaaacttgaagctcattatcagagtccca 1735

Db 1435 GAGGAGAACCTTCTTCGGACCGAAAGAGGAGGTTAAATTAGAGGCTCATATAAAGTGCCA 1494

Qy 1736 tccttctgctgagcagatttattgaaaagagggcaaaacggtgaaatgaacttcagaat 1795

Db 1495 TCATTCTGCTGCGCAGATCATTTGGTAAAGGAGGCAAAACGGTGAACGACTACACAAC 1554

Qy 1796 ttgtcaagtgcagaagtgtgttgcctcgtgacacacacctgatgagaatgaccaaagt 1855

Db 1555 TTGACCAAGTCGGAGGTGGTGGTCCCAACGGGACCAGACACCTGTATGAATAATGACCAAGTA 1614

Qy 1856 gtgtcaaaaactggtcacttctatgttgcaggttccacagagaaaaattcagaaa 1915

Db 1615 GTAGTAAATATCAGAGGCCACTTCTATGCAAGCCAGCTGGCCCAAGAGATGCCAGAG 1674

Qy 1916 atttgcactcaggttaagcagcaccacacagaagctctgcaaaagtggaccacctcag 1975

Db 1675 ATATCTCTCAGGTGAGCGGCAACAGCAGCCCAAGCCCTCAGCCGAGGTCCACCACTC 1734

Qy 1976 tcaagcaggaagtaaaagctcag 1998

Db 1735 GCCAGAGGAATAGAGCCACTG 1757

RESULT 15

AC023375/c

LOCUS

DEFINITION

AC023375

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC023375 158105 bp DNA HTG 08-MAR-2000

Homo sapiens chromosome 8 clone RP11-571M6 map 8, WORKING DRAFT

SEQUENCE, 37 unordered pieces.

AC023375

AC023375.2 GI:7209933

HTG; HTGS_PHASE1; HTGS_DRAFT.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 158105)

Bliren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 8, clone RP11-571M6

Unpublished

2 (bases 1 to 158105)

Bliren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Larcocque, K., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,

TITLE
JOURNAL

COMMENT

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2000 this sequence version replaced gi:6970513.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4227

Center clone name: 571_M_6

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 133946 bases at least Q40

Consensus quality: 143828 bases at least Q30

Consensus quality: 149412 bases at least Q20

Insert size: 174000; agarose-fp

Insert size: 154505; sum-of-contigs

Quality coverage: 3.2 in Q20 bases; agarose-fp

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1036: contig of 1036 bp in length

* 1037 1136: gap of 100 bp

* 1137 2213: contig of 1077 bp in length

* 2214 2313: gap of 100 bp

* 2314 3566: contig of 1253 bp in length

* 3567 3666: gap of 100 bp

* 3667 4804: contig of 1138 bp in length

* 4805 4904: gap of 100 bp

* 4905 6404: contig of 1500 bp in length

* 6405 6504: gap of 100 bp

* 6505 8366: contig of 1862 bp in length

* 8367 8466: gap of 100 bp

* 8467 10126: contig of 1660 bp in length

* 10127 10226: gap of 100 bp

* 10227 11460: contig of 1234 bp in length

* 11461 11560: gap of 100 bp

* 11561 13697: contig of 2137 bp in length

* 13698 13797: gap of 100 bp

* 13798 15334: contig of 1537 bp in length

* 15335 15434: gap of 100 bp

* 15435 16434: contig of 1000 bp in length

* 16435 16534: gap of 100 bp

* 16535 18900: contig of 2365 bp in length

* 18901 19000: gap of 100 bp

* 19001 20918: contig of 1918 bp in length

* 20919 21018: gap of 100 bp

* 21019 23110: contig of 2092 bp in length

* 23111 23210: gap of 100 bp

* 23211 25903: contig of 2693 bp in length

* 25904 26003: gap of 100 bp


```
* 26004 29042: contig of 3039 bp in length
* 29043 29142: gap of 100 bp
* 29143 32196: contig of 3054 bp in length
* 32197 32296: gap of 100 bp
* 32297 35269: contig of 2973 bp in length
* 35270 35369: gap of 100 bp
* 35370 39031: contig of 3662 bp in length
* 39032 39131: gap of 100 bp
* 39132 42030: contig of 2899 bp in length
* 42031 42130: gap of 100 bp
* 42131 46601: contig of 4471 bp in length
* 46602 46701: gap of 100 bp
* 46702 51387: contig of 4886 bp in length
* 51388 51487: gap of 100 bp
* 51488 55227: contig of 3740 bp in length
* 55228 55327: gap of 100 bp
* 55328 59146: contig of 3819 bp in length
* 59147 59246: gap of 100 bp
* 59247 62533: contig of 3293 bp in length
* 62540 62639: gap of 100 bp
* 62640 66443: contig of 3804 bp in length
* 66444 66543: gap of 100 bp
* 66544 72068: contig of 5525 bp in length
* 72069 72168: gap of 100 bp
* 72169 78427: contig of 6259 bp in length
* 78428 78527: gap of 100 bp
* 78528 83366: contig of 4839 bp in length
* 83367 83466: gap of 100 bp
* 83467 89726: contig of 6260 bp in length
* 89727 89826: gap of 100 bp
* 89827 97979: contig of 8153 bp in length
* 97980 98079: gap of 100 bp
* 98080 106168: contig of 8089 bp in length
* 106169 106268: gap of 100 bp
* 106269 114387: contig of 8119 bp in length
* 114388 114877: gap of 100 bp
* 11488 121719: contig of 7232 bp in length
* 121720 121819: gap of 100 bp
* 121820 130088: contig of 8269 bp in length
* 130089 130188: gap of 100 bp
* 130189 139207: contig of 9019 bp in length
* 139208 139307: gap of 100 bp
* 139308 158105: contig of 18798 bp in length.
FEATURES
    source
    1..158105
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /map="8"
    /chromosome="8"
    /clone="RP11-571M6"
    /clone_lib="RPC1-11 Human Male BAC"
misc_feature 1..1036
    /note="assembly_fragment"
misc_feature 1137..2213
    /note="assembly_fragment"
misc_feature 2314..3566
    /note="assembly_fragment"
misc_feature 3667..4804
    /note="assembly_fragment"
misc_feature 4905..6404
    /note="assembly_fragment"
misc_feature 6505..8366
    /note="assembly_fragment"
misc_feature 8467..10126
    /note="assembly_fragment"
misc_feature 10227..11460
    /note="assembly_fragment"
misc_feature 11561..13697
    /note="assembly_fragment"
misc_feature 13798..15334
    /note="assembly_fragment"
misc_feature 15435..16434
    /note="assembly_fragment"
```

```
misc_feature 16535..18900
    /note="assembly_fragment"
misc_feature 19001..20918
    /note="assembly_fragment"
misc_feature 21019..23110
    /note="assembly_fragment"
misc_feature 23211..25903
    /note="assembly_fragment"
misc_feature 26004..29042
    /note="assembly_fragment"
misc_feature 29143..32196
    /note="assembly_fragment"
misc_feature 32297..35269
    /note="assembly_fragment"
misc_feature 35370..39031
    /note="assembly_fragment"
misc_feature 39132..42030
    /note="assembly_fragment"
misc_feature 42131..46601
    /note="assembly_fragment"
misc_feature 46702..51387
    /note="assembly_fragment"
misc_feature 51488..55227
    /note="assembly_fragment"
misc_feature 55328..59146
    /note="assembly_fragment"

Query Match 20.2%; Score 842.8; DB 2; Length 158105;
Best Local Similarity 88.8%; Pred. No. 7.5e-160;
Matches 847; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 1885 ttgccagggtgccagagaaaaattcaggaaattctgactcaggtaaaagcagcaccaca 1944
    |||
Db 9309 TTTCTAGGTGCCCAGAGAAAAATTCAGGAATTCGACTCAGGTAAACAGCACCACA 9250

Qy 1945 acagaaggctctgaaagtggaccacctcagtcagacggaagtaaggctcaggaaaa 2004
    |||
Db 9249 ACAGAAGGCTCTGCAAGTGGACCACCTCAGTCAAGACGGAAGTAAAGGCTCAGGAACA 9190

Qy 2005 gccaccacagaggcagatgccaaaccaaagacagattgcttaaccaacagatggcgct 2064
    |||
Db 9199 GCCCACCACAGGCGAGATGCCAAACCAAGACAGATTGCTTTAACCAACAGATGGCGCT 9130

Qy 2065 gacccctatccagaatcacatgcacaagttttacctagccagttgttctgaggacca 2124
    |||
Db 9129 GACCCCTATCCAGATCAGATGCACAGTTTTTTACCTAGCCAGCTGTTTCTGAGGACCA 9070

Qy 2125 ggcaactttgaaactcctgtctctgtgagaatgtatactttatgctctctgaaatgatg 2184
    |||
Db 9069 GGCAACTTTTGAACCTCTCTGTGTGAGATGTATACCTTTATGCTCTCTGAAATGTATG 9010

Qy 2185 acaccagctttaaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 2244
    |||
Db 9009 ACACCCAGCTTTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 8950

Qy 2245 gagaagagctctgcacttccctttgtgtgagtcacagataacagatatcttaattct 2304
    |||
Db 8949 GAGAAGAGCTCTGCACCTTCCCTTTGTTGTAGTCTACAGTATACAGATATCTAATTCT 8890

Qy 2305 tottaattcccccataatgccagaaaattggcttaatgatgctttcactaaattcatca 2364
    |||
Db 8899 TCTTAATATTCCTCCCATTAATGCCAGAAATTTGGCTTAATGCTTTCACTAAATTCATCA 8830

Qy 2365 aatagattgctcctaataatgccaaattgttaaaattggatcagaataattatcacaggaaatt 2424
    |||
Db 8829 AATAGATTGCTCCTAAATCCAAATTTTAAATTTGGATCAGATAATTAATTAACAGGAATT 8770

Qy 2425 aaattgtaagccattagcatagaaaaactgttctcagtttttttttttttttttttttt 2484
    |||
Db 8769 AAATGTTAAGCCATTAGCATAGAAAAACHTTCTCAGTTTTTATTTTACCTAACATAAC 8710

Qy 2485 atgagtaacctaaaggaagtgctgaatggtgttgagggggtattaaacgttcattttta 2544
    |||
```

[illegible]

Search completed: January 25, 2002, 22:49:36
Job time: 28513 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	4143	99.1	4159	4	US-09-061-709-4	Sequence 4, Appli
2	804.4	19.2	2224	4	US-09-261-855-1	Sequence 1, Appli
3	710.4	17.0	3412	4	US-09-061-709-6	Sequence 6, Appli
4	695.6	16.6	3283	4	US-09-061-709-8	Sequence 8, Appli
5	672.6	16.1	1946	4	US-09-061-709-7	Sequence 7, Appli
6	580.6	13.9	1708	4	US-09-061-709-5	Sequence 5, Appli
7	48.8	1.2	7218	1	US-08-232-463-14	Sequence 14, Appli
8	44.8	1.1	518	1	US-08-485-284A-2	Sequence 2, Appli
9	44.8	1.1	14796	4	US-08-975-080-35	Sequence 35, Appli
10	44.8	1.1	14796	4	US-09-630-706-10	Sequence 10, Appli
11	43.8	1.0	665	2	US-08-883-795A-36	Sequence 36, Appli
12	41.8	1.0	7218	1	US-08-232-463-14	Sequence 14, Appli
13	41.6	1.0	5852	1	US-07-867-106-2	Sequence 2, Appli
14	41.2	1.0	5852	1	US-07-867-106-2	Sequence 2, Appli
15	40.8	1.0	742	1	US-07-847-010-12	Sequence 12, Appli
16	40.8	1.0	1562	1	US-08-330-108-1	Sequence 1, Appli
17	40.8	1.0	1562	5	PCR-US92-10087-1	Sequence 1, Appli
18	40.6	1.0	4159	4	US-09-061-709-4	Sequence 4, Appli
19	40.4	1.0	6243	2	US-09-056-075-1	Sequence 1, Appli
20	40.2	1.0	1758	3	US-08-836-567-3	Sequence 3, Appli
21	39.8	1.0	2726	1	US-08-281-714-1	Sequence 1, Appli
22	39.8	1.0	37950	4	US-09-338-907-183	Sequence 183, Appli
23	39.6	0.9	4140	3	US-08-894-731-2	Sequence 2, Appli
24	38.8	0.9	4235	1	US-08-021-601-3	Sequence 3, Appli
25	38.8	0.9	4235	1	US-08-082-849B-3	Sequence 3, Appli
26	38.8	0.9	4235	5	PCR-US94-01624-3	Sequence 3, Appli
27	38.8	0.9	51952	3	US-08-947-823-1	Sequence 1, Appli

Db 2461 gttttatttttaactaaactaacatgagtaacctaaaggaagtgtgaatggttggc 2520
QY 2521 aggggtattaaacgtgcatttttaactcaactacccacagttatctcagtaatacaatgaataa 2580
Db 2521 aggggtattaaacgtgcatttttaactcaactacccacaggtatctcagtaatacaatgaataa 2580
QY 2581 gcaaaatgttccctttttttttgaaatatttatactatttataatgatagaagtcaccaacg 2640
Db 2581 gcaaaatgttccctttttttttgaaatatttatactatttataatgatagaagtcaccaacg 2640
QY 2641 ttttttaaaaaataaatttaaaatttaacagcaatcaacagctaacagggcaaataaagatttt 2700
Db 2641 ttttttaaaaaataaatttaaaatttaacagcaatcaacagctaacagggcaaataaagatttt 2700
QY 2701 tacttctggctggtgacagtaaaagctggaataatttaatttcaggggttttttgaggttttg 2760
Db 2701 tacttctggctggtgacagtaaaagctggaataatttaatttcaggggttttttgaggttttg 2760
QY 2761 acacagttattagtttaaaatcaaatgttcaaaaatacaggagcagtgctcagtagtctggaga 2820
Db 2761 acacagttattagtttaaaatcaaatgttcaaaaatacaggagcagtgctcagtagtctggaga 2820
QY 2821 gcagcactaccatttattcttccatttatagttggaaaagtttttgacggtactaacaata 2880
Db 2821 gcagcactaccatttattcttccatttatagttggaaaagtttttgacggtactaacaata 2880
QY 2881 gtggtgcagggagattttggaacgctggttttaaatgtgcttcagagacttcagtttttt 2940
Db 2881 gtggtgcagggagattttggaacgctggttttaaatgtgcttcagagacttcagtttttt 2940
QY 2941 gtttagctacatgattgaatgcataataaatgctttgtgcttctgactatcaatacctaa 3000
Db 2941 gtttagctacatgattgaatgcataataaatgctttgtgcttctgactatcaatacctaa 3000
QY 3001 agaaagtgcatcagtgaaagagatgaagacttccaactgactggcaaaaagaacgtttca 3060
Db 3001 agaaagtgcatcagtgaaagagatgaagacttccaactgactggcaaaaagaacgtttca 3060
QY 3061 gctgtcttatagatgcttagtttgccactacactcagaccatggagacagtacaga 3120
Db 3061 gctgtcttatagatgcttagtttgccactacactcagaccatggagacagtacaga 3120
QY 3121 tgggtgacagtggtttaaacgcacaaagagctacatttccatggggccagcactgtcat 3180
Db 3121 tgggtgacagtggtttaaacgcacaaagagctacatttccatggggccagcactgtcat 3180
QY 3181 gagcctcactaagctattttgagatttttaagcactgataaatttaaaaaaataaaaaa 3240
Db 3181 gagcctcactaagctattttgagatttttaagcactgataaatttaaaaaaataaaaaa 3240
QY 3241 aaattagactccaccttaagtagtaaaagtataacaggatttctgtatctgtgcaatcag 3300
Db 3241 aaattagactccaccttaagtagtaaaagtataacaggatttctgtatctgtgcaatcag 3300
QY 3301 ttctttgaaaaaaagtcacaaagatagagaatacaagaataagtttttngggatataatttg 3360
Db 3301 ttctttgaaaaaaagtcacaaagatagagaatacaagaataagtttttngggatataatttg 3360
QY 3361 aatgactgtgaaaacatatgacctttgtataacgaaactaatttgcactccttgacagca 3420
Db 3361 aatgactgtgaaaacatatgacctttgtataacgaaactaatttgcactccttgacagca 3420
QY 3421 aagccagtagctacaattgtttgggtgtgggtgtgggttcccaagccacgctgctctctg 3480
Db 3421 aagccagtagctacaattgtttgggtgtgggtgtgggttcccaagccacgctgctctctg 3480
QY 3481 aattgattttttgagtttttgnntgnaagatgatacacagnatgtttacactatcttnaa 3540
Db 3481 aattgattttttgagtttttgnntgnaagatgatacacagnatgtttacactatcttnaa 3540
QY 3541 ggacatatntataacccttttaaaaaaataccccctgctcctattcttatttcgagatga 3600
Db 3541 ggacatatntataacccttttaaaaaaataccccctgctcctattcttatttcgagatga 3600

QY 3601 atttcgatcacagactagatgtctttctgaagatcaattagacatttntgaaaatgatttaa 3660
Db 3601 atttcgatcacagactagatgtctttctgaagatcaattagacatttntgaaaatgatttaa 3660
QY 3661 agtgttttcttaattgtctctgaaaacaagtcttctttttagtattttaaccacaaaagt 3720
Db 3661 agtgttttcttaattgtctctgaaaacaagtcttctttttagtattttaaccacaaaagt 3720
QY 3721 ccccttttgcactggtttctcctagcattcatgatttttttttcacacaatgaattaaa 3780
Db 3721 ccccttttgcactggtttctcctagcattcatgatttttttttcacacaatgaattaaa 3780
QY 3781 attgctaaaaatcatgactggtcttctggttgcattcaggttaagtggtttlaaggcca 3840
Db 3781 attgctaaaaatcatgactggtcttctggttgcattcaggttaagtggtttlaaggcca 3840
QY 3841 gagcttttctcagttattgatttttttttccccaataattgattttttaaaaaatacacat 3900
Db 3841 gagcttttctcagttattgatttttttttccccaataattgattttttaaaaaatacacat 3900
QY 3901 aggagctgattttaaaccctgctggttttaattctgcattctcanatttctcagccttta 3960
Db 3901 aggagctgattttaaaccctgctggttttaattctgcattctcanatttctcagccttta 3960
QY 3961 gtatggcnaatcanaatttacttttacttaagcatttgaatttggagtatctggtacta 4020
Db 3961 gtatggcnaatcanaatttacttttacttaagcatttgaatttggagtatctggtacta 4020
QY 4021 gtaagaataataatcnaatgagtttcttactcnaaanaatgggtcattctcctcatgn 4080
Db 4021 gtaagaataataatcnaatgagtttcttactcnaaanaatgggtcattctcctcatgn 4080
QY 4081 ataagttnccccaatgcagcttcatttccaganaaccttgacgcaggataaattttttc 4140
Db 4081 ataagttnccccaatgcagcttcatttccaganaaccttgacgcaggataaattttttc 4140
QY 4141 atcatttagtcccccaaaa 4159
Db 4141 atcatttagtcccccaaaa 4159

RESULT 2

US-09-261-855-1
; Sequence 1, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; FILE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296, 95131
; CURRENT APPLICATION NUMBER: US/09/261, 855A
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2224
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-261-855-1

Query Match 19.2%; Score 804.4; DB 4; Length 2224;
Best Local Similarity 67.0%; Pred. No. 6.9e-199;
Matches 1141; Conservative 0; Mismatches 561; Indels 0; Gaps 0;
QY 247 ccaatgaacacactgtatatcgaaacctcagcgagaaacgcgccttcggacactaga 306
Db 127 caccatgaacacagctttacatcggaacctcaacgagagtgtagaccccgacagcttgg 186
QY 307 aagtatcttcaaggacgcacaaagatcccggtgtcgacccttctcgttggaagactggcta 366
Db 187 gaaagtatttcgagcagacaaagatctcctacagcgccaggttcttggtaaaatccggcta 246

```
QY 367 cgcattcgtgactgcgcggagcagagactgggcctccaaaggccatcagggcgcttttcagg 426
DB 247 cgccttcgtgattgcgcgcagcagcactggcgatgaaggccatcgaactttctcggg 306
QY 427 taaatagaactgcacgggaaaccatagaaagtggagcactcggtcccaaaaggcgaag 486
DB 307 gaaagtgaagactgaaggaaacgtctagagattgaaacactcagtcaccaaaacaaag 366
QY 487 gattcgaacactcagatagaaatatccgcctcatttacagtgggagggtgctgggatag 546
DB 367 gagtcgaaatacagatccgcaatatccacctcagctccgatgggaagtgcctagatag 426
QY 547 ttactagtcagatggagtggggagagctgtgagcaagtgaacactgactcggaaac 606
DB 427 cctgctggctcagtcacgtggaagactgtgagcaagtgaacactgaaagtggagac 486
QY 607 tgcagtgtaaaatgaacctattccagtaaggacacagctagacacactagacaact 666
DB 487 agcgttggaacgtcaccttactctaacccgggagcagaccaggcaagctatcatgaagct 546
QY 667 gaattgattcagtagagaaatttcaccttgaaagttagcctatatccctgatgaaatggc 726
DB 547 aaatggccatcaactggagaacatgcccgaaggtctctacatacctgatgagcagat 606
QY 727 cgcacagaaaccccttcagcagaccccagaggtcgccggggccttgggcagaggggctc 786
DB 607 aacacaaggctcctgagaaatggcgctgtgaggctttgggtctcggggccagccccgca 666
QY 787 ctcaaggcaggggtctccagatccgtatccaagcagaacacatgtatttgcctctgag 846
DB 667 aggtctcccgctggcagcaggggtccagccaagcagcagccagtgacatccctctccg 726
QY 847 cctgctggttcccaaccatattgttgagccatcataggaaagaaaggtgcaccacttcg 906
DB 727 gctcctggtgctacgcagtatgagcgctatcatctatggcaaggagggtgccaccatccg 786
QY 907 gaaacatcaccaaacagaccagcttaaaatcgatgtccaccgtaaagaaagaggggggc 966
DB 787 aaacatcacaaacagacgcagctcaaaatagacgtgcataggaaggagagatggcggc 846
QY 967 tgcagaaagtcgattactctctactctcctcctgagggcacctctcgcttgtaagtc 1026
DB 847 tgcggaagggccatcagctgcatcaacccctgaaaggctgctccctccgcgcaagat 906
QY 1027 tattctggagattatgcataaggaagctcaagatatataaaattcacagaagagatccctt 1086
DB 907 gatcttgagattatgcagaaggaggcaaggacacaaacggcagatgaagtccct 966
QY 1087 gaagattttagctcataataaactttgttgagctcttattgttgtaagaaaggaatact 1146
DB 967 gaagatcctggctcataacaactctcgtcggcgactcattggcaaggagcgggaacct 1026
QY 1147 taaaaaatlgagcaagacacacactaaaatcacgatatctcattcgaggaaattgac 1206
DB 1027 gaagaaagtgaggagacacagacagagagatcaacatctcatcgtccagacacctac 1086
QY 1207 gctgataatccgaacgcactattacagttaaaggcaatgttgagacatgtgccaagc 1266
DB 1087 gctctaaacctgagaggaccatcactgtgaaggcgccatctgagaactgttcagggc 1146
QY 1267 tgaggagagatcatgaagaaatcaggggaggtttatgaaaaatgatatgtctctatgaa 1326
DB 1147 cgagcaggagatcatgaagaaagtcgagaggtttagaagacgacgtggccgcctag 1206
QY 1327 tctcaagcagcatattactctgattaaatctgaaacgcttggtctgtgtccaccac 1386
DB 1207 ctgagctccaccctatccctggcttaacctggctgctgtgaggtctctccagcttc 1266
QY 1387 ttcaggatgccacctccaccctcagggcccccttcagccatgactcctccctaccgcga 1446
DB 1267 atccagcgctgtccctctcctcccagcagtgctactgggggtgctccctctatgctcctt 1326
```

```
QY 1447 gttgagcaatcagaacacgagagactgttcatcagtttatccaccagctctatcagtcgc 1506
DB 1327 catgaggtccggagcaggagatggtacaagtgttcatcccccaggtgtgggggc 1386
QY 1507 catcatcggaacgagcggccagacacatcaagcagcttctcgtttcttgagactcaat 1566
DB 1387 catcattggcaagaggccagcacatacaaaacactctcccgtttcccgagcgtccat 1446
QY 1567 taagattgctccagcggaagcaccagatgctaaagtggaggtggtattatcactggacc 1626
DB 1447 caagattgctccaccagaaacacactgactccaaagttcgaagtgtcgtcatcactggacc 1506
QY 1627 accaaggtcagttcaaggtcagggaggaatattgaaaaattaaagaagaaaactt 1686
DB 1507 cccagaggtcagttcaaggtcagggaggaatatttggcaaaactaaagaagaaattt 1566
QY 1687 tgttagtccctaaagaagaggtgaaactgaagctcatatcatcagagtgccatcttgcgc 1746
DB 1567 ctttgggtcccaaggaggaaagtaaaagctagagaccatatacgggttccgggttcagcgc 1626
QY 1747 tggcagagttattgaaaaaggagcgaacgggtgaatgaacttcagaattttgtcaagtgc 1806
DB 1627 cggcggctcatcggaaggcgcaaaagcggtaaatgagctgcagaacttgactgcgc 1686
QY 1807 agaagtgtgtcctcgtgaccagacacctgtatgagaatgaccaagtgtgtgtcaaaat 1866
DB 1687 tgaggtggtatgccaagagaccagacccggatgagaagaccaagtcattgttaagat 1746
QY 1867 aactggtcacttctatctgttcagaggttgcagagaaaaattcagaaaattcgaactca 1926
DB 1747 catgggacattctatgccagccagatgggtcagcggaaagatcccgagacatccctggctca 1806
QY 1927 ggtaagcagcaccaacaacag 1948
DB 1807 agttaagcaacagcaccagaag 1828

RESULT 3
US-09-061-709-6
; Sequence 6, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-6

Query Match 17.0%; Score 710.4; DB 4; Length 3412;
Best Local Similarity 64.5%; Pred. No. 2.1e-174;
Matches 1155; Conservative 0; Mismatches 556; Indels 81; Gaps 3;

QY 251 atgaacaactgtatatcggaacactcagcagagaacgcgcgcctcgacacctagaagt 310
DB 73 atgaacaagctttatcatcggaacactgagccgcgcctcaccgcgcgacacctcgcgag 132
QY 311 attctcaagcagcccaagatcccggtgttcgggacctctcgtgtgaagactggtctacgcg 370
```

Db 133 ctctttggggacaggaagctgccctggcgggagacaggctcctgctgaagtcocggtaagcc 192
QY 371 ttctggactgccgagagagctgggcccctcaaggccatcgagcgctttcaggtaaa 430
Db 193 ttctgtgactaccgccaccagaaactgggcccattccggccatcgagacctctcgggtaaa 252
QY 431 atagaactgcaagggaacccataagaattgagcactcgggtcccaaaaagcaaggatt 490
Db 253 ggggaattgcatgggaaaaatcaggaagtgtattactcagctctaaaaagctaaaggagc 312
QY 491 cggaaacttcagatcagaaaaatccgctcatttaccagtggaggtgctgtagattta 550
Db 313 aggaaaaattcagatttcgaaaactccctccaccgcgagtggaggtgttgatggact 372
QY 551 ctagtccagtatggagtggtagagctgtgagcaagtgaacactgactcggaaactgca 610
Db 373 ttggctcaatatggacagtggagaatgtggaacaagtcaacacagacagaaaaccgcc 432
QY 611 gtgttaaatgtaacctattccctagtaagacacagctagacaagcactagacaaactgaat 670
Db 433 gtgtcaactgcacatatgcaacaagagaagaagcaaaaatagccatggagaagctaaagc 492
QY 671 ggatttcagttagaaatttcaaccttgaaagttagccttatccctgatgaaatggccgcc 730
Db 493 gggccatcagtttgagaactactccctccaagatttctcatatcccggatgaagagtgaagc 552
QY 731 cagcaaaaaccccttcgacagcccgcgaggtgcgcggggcttgggcagaggggctccctca 790
Db 553 tcccttcgccccctcagcgagccagcagcggtggggaccactctccgggagcaaggccac 612
QY 791 aggcagggggtctccaggatccgtatccaaagcagaaccatgtgattgctcctgcgcctg 850
Db 613 g-----cccttggggccacttctcagggccagacagattgatttcccgctgcggatc 663
QY 851 ctgggtccaccacaaattgttgagccatcatagaaaagaagggtgcacacattcgggaac 910
Db 664 ctgggtccaccacaaattgttggtgccatcatcgaaaaggaggggttgaccataaagaac 723
QY 911 atcaccaaacagaccagctctaaaatcgatgtcccacgttaaaagaaaatgcggggctgct 970
Db 724 atcataagcagaccagctcccggtagatataatccatagaaaaagagaactcttgagctgca 783
QY 971 gagaagtcgattactctctactcctgaagcacctctgcggctgtgtaagtctatt 1030
Db 784 gagaagctgtcacctatccatccaccocagagggactctgaagcatgcgcgatgatt 843
QY 1031 ctggagattatgataaaggaaactcaagataataaattcaagaagagatccoccttgaag 1090
Db 844 ctggaatcatgcaaaaagaggcagatgagaccaaactagccgaagagattcctctgaaa 903
QY 1091 attttagctcataataactttgttgacgtcttattggtaaaagagggaagaatctttaa 1150
Db 904 actctggcacacaatggctgtgttggaagactgattggaaaaagaggcagaaaatttgaag 963
QY 1151 aaaattgagcagacacagacactaaaatcacgatatctccatttcaggaaattgacgtg 1210
Db 964 aaaattgacatgaacaggggcccagaagataacaactctatctcttgcaggatttgacata 1023
QY 1211 tataatccagaacgcactattacagttaaaggcaatgttgagacatgtgccaaaactgag 1270
Db 1024 tacaacccgggaagaaccatcactgtgaaggggcagcagttgagggcgtgtgccagtgctgag 1083
QY 1271 gaggagatcataaagaaatccggaggtcttatagaataatgatatgtcttctatgaatct 1330
Db 1084 atagagattatgaagaagctgcgtgagggcctttgaaaatgatatgtcgtggtttaaccaa 1143
QY 1331 caagcacatttaattcctggattaaatctgaacgccttgggtctgttcccaaccacttca 1390
Db 1144 caagcacaattctgatacccagggttgaaacctcagcgcacttggcatcttttcaacagagctg 1203
QY 1391 gggatgccacctccacctcagggccccccttcagccatgaactcct----- 1435
Db 1204 tccgtgctatctccaccagaggcccccggagctcccccgcgtgccccctaccacccc 1263

QY 1436 -----ccctaaccgcagtttgagc----- 1454
Db 1264 tttaactaaccactccoggatacttctccagcctgtaccctccatcacccagtttggccggttc 1323
QY 1455 -----aatcagaaaacggagactgttcatcagtttccacagcttatccacagcttatca 1498
Db 1324 ccgcatcatcactcttccagagcaggagattgtgaactcttctatcccaaccaggct 1383
QY 1499 gtccgtggccatcatcggcgaagcaggccagcacatcaagcagcttctcgtcttgcgtgga 1558
Db 1384 gtgggcgcctatcatcgggaagggggcacacatcaaacagctgcggagattcgcggga 1443
QY 1559 gcttcaattaaagtctccagcgggaagcaccagctgctaaagtggaggtggtgattatc 1618
Db 1444 gcctctatcaagatgccctcgcggaagcccgagcgtcagcgaaaaggatggtcatcatc 1503
QY 1619 actgaaacccagaggtcagttcaaggctcagggggaagaattttatggaaaaataaagaa 1678
Db 1504 accgggccaccggaagccagttcaaggccagggccagcagctcagcgaaaaggatggtcatcatc 1563
QY 1679 gaaaactttgttagtctctaaagaagaggtgaaacttgaaactgaagctcatatcagagtgcctatcc 1738
Db 1564 gaaaactctttaaaccctcaagaagaagtggaagctggaagcgcataatcagagtgccctct 1623
QY 1739 ttgtgctggcagagatttattgaaaaaggaggcaaacgcgtgtaataacttcagaatttg 1798
Db 1624 tccacagctggcgggtgattggcaaaagggtggcaagacgtgaaacgaactgcagaactta 1683
QY 1799 tcaagtgcgaagttgttgcctcctgcacagacacacctgatgagaatgaccaaagtggtt 1858
Db 1684 accagtgcaagaatcatcgtgcctcgtgaccaaacccagatgaaaatgagggaagtgc 1743
QY 1859 gtcaaaaataactggctcacttctatccttccaggttgcacagagaaaaatttcaggaaatt 1918
Db 1744 gtcaaatattcggcactcttcttctgtagcactgcacagcgcgaagatcaggggaaatt 1803
QY 1919 ctgactcaggtgaagcagcaccacaacacgaagagctctgcacaaagtggaccac 1970
Db 1804 gtacaacaggtgaagcagcagagcagaataaccttcaggggagtcgcctcac 1855

RESULT 4
US-09-061-709-8
; Sequence 8, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 8
; LENGTH: 3283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-8

Query Match 16.6%; Score 695.6; DB 4; Length 3283;
Best Local Similarity 64.8%; Pred No. 1.4e-170;
Matches 1115; Conservative 0; Mismatches 539; Indels 66; Gaps 3;
QY 251 atgaacaaactgtatatcggaacctcagcgagaagcgcgcgcctcgcgacctagaaagt 310

```
Db 73 atgaacagctttacatcgaggaaacctgagccccccgtcacccgcgacacctccgcag 132
Qy 311 atcttaaggagcccaagatccccggttcggaacccctctctggtgaagactggctacgcg 370
Db 133 ctctttgggacaggaagctgccccctggcggaagctctctgctgaagtcgcgctacgcc 192
Qy 371 ttcgtgactgcccggacgagagctgggcccctcaagccatcgagggcgcttccaggtaaa 430
Db 193 ttcgtgactaccccgcacgaagctggccatccgcgcacatcgagacctctcggtgtaa 252
Qy 431 atagaactgcacgggaacccatagaagttgagcactcgctcccaaaaagcaagatt 490
Db 253 gtgaattgcattgggaataatcatggagttgattcactcaagctctcaaaaagctaaagagc 312
Qy 491 cggaaacttcagatcacgaataatccccctcattttacagtgggaggtgctggaagtta 550
Db 313 aggaataatccagattcgaataatccccctcactcactgcagtgggaggtgttgatggactt 372
Qy 551 ctagtccagtatgagtggtgagagctgtgagcaagtgaacactgactcggaactgca 610
Db 373 ttggtcaaatgggacagtggaagtgtgaaacaagtcaacacagacacagaaacgcc 432
Qy 611 gttgtaaatgaacctattccagtaagacacaagctagacaagcactagacaaactgaat 670
Db 433 gttgtcaactgcacatgtcaacaaggagaagcaaaaatagccaagggaagctaaagc 492
Qy 671 ggatttcagttagagaatttcaccttgaaagtagccttatccctgatgaaatggcgcc 730
Db 493 gggcatcagtttgagaactactctcacaagatttccatcacatcccgatgaagagtgagc 552
Qy 731 cagcaaaaaccccttgacgagcccgaggtgcgggggttggcgagagggggtccctca 790
Db 553 tccctctgcgccctcagcagagcccgagcgtgggacacactcttcccgggagcaagc --- 609
Qy 791 aggcaggggtctccagatcgatccaaagcagaaacatgtgaatttgcctctgcgctg 850
Db 610 -----cagccccctgggcaactctcagccagacagattgatttcccgctgcgagtc 663
Qy 851 ctggttcccccccaatttggtagccatcataggaagaaagaggtgcacacttcggaac 910
Db 664 ctggttcccccccaatttggtagccatcatcggaagagggcttgaccataaagaac 723
Qy 911 atcaccaaaacagaccagctctaaatcgatgtccaccgttaaaagaaatgcggggctgct 970
Db 724 atcaatgaagcaccagctcccggttagatcatcatagaaagaaagaaactctggagctgca 783
Qy 971 gagaagtcgattactactctactctcctgaagcaccctctgcgctgtgaagctatt 1030
Db 784 gagaagcctgtcaccatccatgccaccgcaggggggactctctgaagcatgcgcgatgatt 843
Qy 1031 ctggagattatgcataaggagctcaagatatataaaattcacagaagagatccccctggaag 1090
Db 844 ctggaatcatgcagaagagagcagatgagaccaaactagccgaagagattctctgaaa 903
Qy 1091 attttagctcataataactttgttgagctcttatttggtaaaagaaagaaatctttaa 1150
Db 904 atcttgcacacaaatgcttgggtggaagctgattggaagaaagaaagcagaaattgaa 963
Qy 1151 aaaaatgagaacacagacactaaaatcacgatatctccattgcaggaattgagcgtg 1210
Db 964 aaaaatgaaatgaacaggggacaaagataacaatctcatcttgcaggattgagcata 1023
Qy 1211 tataatcagaacgcactattacagtttaaaggcaaatgtgagacatgtgcgaagctgag 1270
Db 1024 tacaaccggaagaacacatcactgtgaaggcagacgttgaggccctgtgcagtgctgag 1083
Qy 1271 gaggagatcatgaagaaatcagggagctcttatgaaatgatattgcttctatgaattctt 1330
Db 1084 atagagattatgaagagctgcgtgagcccttgaaatgatattgctgctgttaacacc 1143
Qy 1331 caagcacatttaattcttgattaaatctgaacgccttgggtgtgttccccaccacttca 1390
```

```
Db 1144 cactc-----cggatacttctccagcctgttacccccataccca 1181
Qy 1391 gggatgcacactccacactcagggcccccttccagcatgaactcctccataccgcagttt 1450
Db 1182 gttggccggtcccgcatcatcactctatc----- 1213
Qy 1451 gagcaatcagaacacggagactgttcatcagtttatccagctctctatcagtcggtycactc 1510
Db 1214 -----cagagcaggagattgtgaatctcttcatcccaaccaggtgtgtggcgccatc 1266
Qy 1511 atcggcaagcaggccacacatcaacagcacttctcgttttctgagcttcaattaaag 1570
Db 1267 atcgggaagagggggcacacatcaacagctggcgagattcgcgcgagcctctatcaag 1326
Qy 1571 attgctcagcggagcaccagatgctaaagtggagtggtgattatcactggaccacca 1630
Db 1327 attgccccctgggaagcccgacgctcagcgaaagagatggtcatcatcacccggccaccg 1386
Qy 1631 gaggctcagttcaaggctcagggaagaatttatggaataatttaaaagaaaaatttggtt 1690
Db 1387 gaagccagttcaaggccccagggagcgtatcttggaaactgaaagagagaaactctttt 1446
Qy 1691 agtctaaagaagaggtgaaacttgaagctcatcatcagagtgccatcttctgctgctg 1750
Db 1447 aaccccaagaagaagtggaagctggaagcgcatacatcagagtgccccctctccacagctggc 1506
Qy 1751 agagtatttggaaagggagggcaaacggtgaaatgaacttcagaatttgcagtgcagaa 1810
Db 1507 cgggtgattggcaaggtggcaagccgtgaacgaactgagaacttaaacagtgcaagaa 1566
Qy 1811 gttgttctccctctgaccagacacctgatgagaatgacaaagtggttgttcaaaaact 1870
Db 1567 gtcatctgctctgaccacacccagatgaaatgagggaagtgcgtcagaattatc 1626
Qy 1871 ggtcactctatgcttgcagggttgcagagaaataattcaggaattctgactcaggtta 1930
Db 1627 gggcattcttctgtagccagactgcacgcgcaagatcaggggaaattgtacaacaggtg 1686
Qy 1931 aagcagcaccaacaacaggaaggctctgcaaaagtggaccac 1970
Db 1687 aagcagcaggagcagaataacctcaggggagtcgctcac 1726

RESULT 5
US-09-061-709-7
; Sequence 7, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Rao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-7
```

Query Match 16.1%; Score 672.6; DB 4; Length 1946;
Best Local Similarity 66.6%; Pred. No. 9.9e-165;
Matches 979; Conservative 0; Mismatches 484; Indels 6; Gaps 1;

Qy 480 ggcacagattcggaacttcagatacgaataatccccgctcattacagtgagggtgc 539
Db 103 gggttagagcggaataatcaatccgaataatccaccccgctccgatgggaagtac 162
Qy 540 tggatattactagtcagatagagtggtgagagactgtgacaagtgaacactgact 599
Db 163 tggacagcctgtgctcagtgagtggttagagagactgtgacaagtgaacaccgaga 222
Qy 600 cggaaactcgagttgtaaaattacactattccagtaggagcgaacagcagcactag 659
Db 223 gtgagacgagtggtgaattgcacctattccaacgggagcagacacgcaagccatca 282
Qy 560 aaaaactgaatgatttcagtagagaaatttcaaccttgaagtgcctatccctgatg 719
Db 283 tgaagctgaattggccaccagttgagagaacctgcccgtgaaggtctctacatccccgatg 342
Qy 720 aaatggccgcccagcaaaaccccttgacgagccccgaggtcgccggggttgaggaga 779
Db 343 agcagatagcacaggggacctgagaattggcgccgaggggttggctctcctgggtcagc 402
Qy 780 ggggtcctcaaggcaggggtctccaggatccgtatccaagcagaaacctgtgatttgc 839
Db 403 ccgcagaggtcacctgtgagcagggggcccccagcagcagcagcaagtggacatcc 462
Qy 840 ctctgcgctgtggtttccaccaccaatttgttgagccatcataggaagaaggtgcca 899
Db 463 cctctggctcctgtgtccaccaccagtatgtgggtgccattattgccaaggaggggcca 522
Qy 900 ccattcggagacatcacaaacagaccagctcctaaatcgatgtccaccgttaaaagaatg 959
Db 523 ccatcccgacacatcacaaacagaccagctcctaaatcgatgtcagtgataggaagagacg 582
Qy 960 cgggggtgctgagaagtgattactactctctctactcctgaagcgacctctcgctt 1019
Db 583 caggtgcagtgaaaaagccatcagtgatgcactccaccctgaggtgtctctccctt 642
Qy 1020 gtaagtctatctggagattatgcataagaagctcaagatatataaaattcacagaaga 1079
Db 643 gtaagatctctggagattatgcataagaagctcaagatatataaaattcacagaaga 702
Qy 1080 tcccttgagagattttagctcataataactttgttgacgtcttatttggttaaaagaaga 1139
Db 703 tccccctgaagatcctggccataataactttgtaggcgctctacttggcaaggaagac 762
Qy 1140 gaaactctaaaaaattgacgaagcacagacactaaataacagatatctccattgcag 1199
Db 763 ggaacctgaagaagtagacaagataccgagacaaataaccatcctctctgtgcaag 822
Qy 1200 aattgcgctgtataatccagaaacgactattacagttaaaggcaatgttgagacatgtg 1259
Db 823 acctaccctttacaacctgagaggaccatcactgtgaaggggggccatcgagaattgtt 882
Qy 1260 ccaaaagctgaggagagatcatgaagaaatcaggaggtcttattgaaatgatattgctt 1319
Db 883 gcaggggccgagcgaataatgaagaagttcgggggacctatgagaatgatgtgctg 942
Qy 1320 ctatgaatcttcaagcacatttaattcctgatttaactatcctgaagccttgggtctgtcc 1379
Db 943 ccataga-----gctctcaactgatccctggcctgaacctggctgtgtgaggtttttcc 996
Qy 1380 caccacttcagggtgaccacctccccactcagggcccccttcagccatgactcctcct 1439
Db 997 cagcttcatccagcgagtcgccgcgctcccagcagcgttacttgggtgctcctcata 1056
Qy 1440 acccgagtttagcaatcagaacacgagagactgttccatcagtttatccagctctatcag 1499
Db 1057 gctcctttatgcagctcccccagcagcagagagtggtgcaggtgtttatccccccagcag 1116
Qy 1500 tcggtgcatcatcggaacagggccagcacatcacagcagcttctcgtcttgcctgag 1559
Db 1117 tgggcccacatcatcggaagaaggggagcacatcaaacagctctcccggttggccagcg 1176
Qy 1560 ctcaataaagattgtctccagcggaagcacagatgctaaagtgaaggtgattatca 1619

Db 1177 cctccatcaagattgcaccaccgaaacacactgactccaaagtctcgtaggttatcatca 1236
Qy 1620 ctgacacacagaggtcagttcaaggtcaggtcaggaagattatgaaaaattaaagaag 1679
Db 1237 ctgagaccgcccagggcccaattcaaggtcaggtcaggaagattctatggaacctcaaggagg 1296
Qy 1680 aaaaactttgttagtctctaaagaaggtgaaacttgaagctcattcagagtgccatcct 1739
Db 1297 agaactcttgggtcccaaggaagaagtgaaagctggagaccacacatacgttgccagcat 1356
Qy 1740 ttgctgtcggcagagttatttgaagaaggaagcgaacacgggtgaattgaactcagaatttgt 1799
Db 1357 cagcagctggccgggtcatttggcaaggtggaaaaacggtgaacgagttgcagaatttga 1416
Qy 1800 caagtgcagaagttgtgtccctcgtgaccagacacctgatgagaaatgacaaagtgtgtg 1859
Db 1417 cggcagctgaggtgtgtagcaccagagaccagaccctcgtatgagaaagaccaggctcag 1476
Qy 1860 tcaaaataactggctcacttctatgcttgcagggttgcacagagaaaaattcagggaatttc 1919
Db 1477 tgaataatcatcggacatttctatgccagtcagatggtctcaacggaagatccgagacatcc 1536
Qy 1920 tgactcaggttaaaagcagcaccacaacacag 1948
Db 1537 tggccaggtttaagcagcagcatcagaag 1565

RESULT 6
US-09-061-709-5
; Sequence 5, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; FILE REFERENCE: LUD 538
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 5
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-5

Query Match 13.9%; Score 580.6; DB 4; Length 1708;
Best Local Similarity 66.1%; Pred. No. 6.8e-141;
Matches 857; Conservative 0; Mismatches 434; Indels 6; Gaps 1;

Qy 652 agcactgacaaaactgaatgatttcagtttagagaatttcacctgaaagtagcctatat 711
Db 37 agccatcatgaagctgaatggccaccagttgagaacctgcccctgaaggtctctacat 96
Qy 712 cccctgatgaatggccgccagcaaaaccccttgcagcagccccgaggtcgccggggct 771
Db 97 ccccgtagcagatagcacaggggacctgagaatgggcgcggaggggcttggctctcg 156
Qy 772 tggcgagaggggctcctcaaggcaggggtctccaggatccgtatccagcagaaacctg 831
Db 157 gggtcagcccccgagggctcacctcactgtgagcggggccccccagcagcagcaagt 216
Qy 832 tgattgctcctgcgctgtggtttcccccacccaaatttggagagccatcataggaaga 891
Db 217 ggacatcccccttggtcctcgttggtgccacccagtatgtgggtgacctatttggcaagg 276


```
Best Local Similarity 54.2%; Pred. No. 0.23; Mismatches 0; Gaps 0;
Matches 91; Conservative 0; Indels 77;

QY 2546 tcaactacctcaggtattcagtaatacaatgaaagcaaaatgttctctttttttgaaa 2605
    || || || || || || || || || || || || || || || || || || || ||
Db 10850 TTATATATCCTAAGTTTACCTCATGTATAGTAATAAATAATTTGTCATTTTAAATATA 10791

QY 2606 attttatatacttataatgatagaagtcacacgctttttttaaataataaatt 2665
    || || || || || || || || || || || || || || || || || || || ||
Db 10790 AATTTATAATTTTAAATTTTAAAGAAAAATTTAAATTTAAAAATAATAATTGATAAGA 10731

QY 2666 taacagaatcagtaacaggaacaaattaagatttttacttctgctgg 2713
    || || || || || || || || || || || || || || || || || || || ||
Db 10730 AAAAGAAAAACAGTGCACCGTTGGTGTGTTTTTTTTTTTAGACACAGG 10683

RESULT 10
US-09-630-706-10/C
; Sequence 10, Application US/09630706
; Patent No. 6277640
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF HER-3 EXPRESSION
; FILE REFERENCE: RTS-0053
; CURRENT APPLICATION NUMBER: US/09/630,706
; CURRENT FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 94
; SEQ ID NO 10
; LENGTH: 14796
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2811)...(2921)
; NAME/KEY: CDS
; LOCATION: (3174)...(3283)
; NAME/KEY: CDS
; LOCATION: (5158)...(5275)
; NAME/KEY: CDS
; LOCATION: (11955)...(12044)
US-09-630-706-10

Query Match 1.1%; Score 44.8; DB 4; Length 14796;
Best Local Similarity 54.2%; Pred. No. 0.23;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 2546 tcaactacctcaggtattcagtaatacaatgaaagcaaaatgttctctttttgaaa 2605
    || || || || || || || || || || || || || || || || || || || ||
Db 10850 TTATATATCCTAAGTTTACCTCATGTATAGTAATAAATAATTTGTCATTTTAAATATA 10791

QY 2606 attttatatacttataatgatagaagtcacacgctttttttaaataataaatt 2665
    || || || || || || || || || || || || || || || || || || || ||
Db 10790 AATTTATAATTTTAAATTTTAAAGAAAAATTTAAATTTAAAAATAATAATTGATAAGA 10731

QY 2666 taacagaatcagtaacaggaacaaattaagatttttacttctgctgg 2713
    || || || || || || || || || || || || || || || || || || || ||
Db 10730 AAAAGAAAAACAGTGCACCGTTGGTGTGTTTTTTTTTTTAGACACAGG 10683

RESULT 11
US-08-883-795A-36/c
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSES: BERESKIN & PARR
```

```
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-883-795A-36

Query Match 1.0%; Score 43.8; DB 2; Length 665;
Best Local Similarity 45.8%; Pred. No. 0.071;
Matches 141; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 3756 tttttttcacacaatgaattgctaaaaatgcataatcatggactggcttctcgttgat 3815
    || || || || || || || || || || || || || || || || || || || ||
Db 334 TTTAATTATAAATATGTAATTTATAACATTTTAAATATAAATAATTTAATTATAAACAT 275

QY 3816 ttcagtgatgtgtttaagccagagctttctcagtgatttcttccccaata 3875
    || || || || || || || || || || || || || || || || || || || ||
Db 274 TTTAATTATAAATATTTAATTTATAAATATTTAATTTATAAATAATTTAATTATAAATAT 215

QY 3876 ttggattttttaaaaatatacacataggagcgtcatttaaaacccctgcgtgttttaattct 3935
    || || || || || || || || || || || || || || || || || || || ||
Db 214 TTTAATTATAAATATTTAATTTATAAATATTTAATTTATAAATAATTTAATTATAAATAT 155

QY 3936 gtcanatttcactctagcccttttagtatggcnaaatcnaaatttactttactlaagcat 3995
    || || || || || || || || || || || || || || || || || || || ||
Db 154 TTTAATTATAAATATTTAATTTATAAATATTTAATTTATAAATAATTTAATTATAAATAT 95

QY 3996 ttgtaattggagtatctgtagctagcagaataaattcnaataattgagttttgtactc 4055
    || || || || || || || || || || || || || || || || || || || ||
Db 94 TTTAATTATAAATATTTAATTTATAAATATTTAATTTATAAATAATTTAATTATAAATAA 35

QY 4056 nccaaanat 4064
    || || || || || || || || || || || || || || || || || || || ||
Db 34 TTTAATTAT 26

RESULT 12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
```


Db	2170	TTTGGCAGTACAAATTTTAATATATATTTATTTTTCGATTAATTTTTTTTTTTTTTTTTTTTTTTTTTTT	2229
Qy	3822	taagatgctgtttaagccagagcctttctcagtatctgattt- ---tttccccaataattt	3878
Db	2230	TAAAAATTTCTTTTTTTTTTTTTTTTTTTTAAATTTTAAATTTTAAATTTTAAATTTTCCACACTTT	2289
Qy	3879	gattttttaaaaatacacacatagcagcgcattttaaaccctgcgtggtttaaattctgtc	3938
Db	2290	CATTTTATTTTATTTTATTTATTTGTAATTCATTTTATTTTATTTTAAATTAATAGATTTT	2349
Qy	3939	anatttcactctagcctcttagtagtggcnaatcnaatttacttttacttacttaagcatttg	3998
Db	2350	GGTTTAATTTTATTCAAAGATTTTAAAATGGCAGCACTTATTTCTTGGGATAGGTTTTT	2409
Qy	3999	taattt 4004	
Db	2410	TAACTT 2415	

```

RESULT 14
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rlis
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

```

```

Query Match          1.0%; Score 41.2; DB 1; Length 5852;
Best Local Similarity 50.5%; Pred. No. 1.2;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps

Qy 3144 acaaaggtacatttcctatggggccagcactgtcatgagcctcactaagctatttgaa 3203
      ||||| || || || || || || || || || || || || || || || || || || ||
Db 2421 ACRAAAACCTTAAAAAACCTATCCCAAGAAATAAGTCGTGCATTTTAAAAATCTTTGAA 2362
      ||||| || || || || || || || || || || || || || || || || || || ||

Qy 3204 gattttaagcactgataattaaaaaataaaaaaataaaataaaataaaataaaataaaataaa 3263
      ||||| || || || || || || || || || || || || || || || || || || ||
Db 2361 TAAATTTAAACCAAAACTATTTTAAATTAATAAAATAAAATGAATTTACAAATAAATAAAA 2302
      ||||| || || || || || || || || || || || || || || || || || || ||

Qy 3264 taaagtataacagatttctgtatctgctgcaatcagttctttgaaaaaaagtcacaaag 3323
      ||||| || || || || || || || || || || || || || || || || || || ||
Db 2301 TAAATTAATAAGAACTGTGGGAAATAAAATTTAAAAATTTAAAAATTAATAAAATAAAAA 2242
      ||||| || || || || || || || || || || || || || || || || || || ||

Qy 3324 atagagaatacagaagaaa 3341
      ||||| || || || || || || || || || || || || || || || || || || ||
Db 2241 AAAGAAATTTTAAAAAAA 2224
      ||||| || || || || || || || || || || || || || || || || || || ||

RESULT 15
US-07-847-010-12
; Sequence 12, Application US/07847010
; Patent No. 5693495
; GENERAL INFORMATION:
; APPLICANT: Breiteneder, Heimo
; APPLICANT: Reikerstorfer, Arnold
; APPLICANT: Valenta, Rudolf
; APPLICANT: Hoffmann - Sommergruber, Karin
; APPLICANT: Breitenbach, Michael
; APPLICANT: Kraft, Dietrich
; APPLICANT: Rumpold, Helmut
; APPLICANT: Scheiner, Otto
; APPLICANT: Ebner, Christof
; APPLICANT: Ferreira, Fatima
; TITLE OF INVENTION: Allergens of Alder Pollen and
; TITLE OF INVENTION: Applications Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE:.. New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/847,010
; FILING DATE: 01-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jones III, Harry C
; REGISTRATION NUMBER: 20,280
; REFERENCE/DOCKET NUMBER: 6530-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 742 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:28 ; Search time 340.59 Seconds

(without alignments)

10524.316 Million cell updates/sec

Title:

US-09-685-696-175

Perfect score:

4181

Sequence:

1 ggtggatgcgttgggtgtg.....aaaaaaaaaaaaaaaa 4181

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched:

930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4165	99.6	4181	21 AAC65900	Human lung cancer-
2	4143	99.1	4159	21 AAZ36150	DNA encoding cance
3	1736.8	41.5	1740	21 AAC66035	Human lung cancer-
4	804.4	19.2	2224	20 AAZ10617	CDNA encoding a mu
5	710.4	17.0	3412	21 AAZ36152	DNA encoding cance
6	695.6	16.6	3283	21 AAZ36154	An alternative for
7	672.6	16.1	1946	21 AAZ36153	An alternative for
8	580.6	13.9	1708	21 AAZ36151	DNA encoding cance
9	409.6	9.8	583	20 AAZ24600	Human lung tumor a
10	409.6	9.8	583	21 AAC65839	Human lung cancer-
11	381.4	9.1	444	21 AAC17226	Human secreted pro

12	282	6.7	282	22	AAI56378	Probe #25064 used
13	282	6.7	588	22	AAI43398	Probe #12084 used
14	263.2	6.3	710	21	AAA02565	Human colon cancer
15	171.2	4.1	936	22	AAF58252	Oligonucleotide D1
16	171.2	4.1	936	22	AAF58254	Oligonucleotide D1
17	171.2	4.1	936	22	AAF58257	Oligonucleotide D1
18	171.2	4.1	936	22	AAF58259	Oligonucleotide D2
19	171.2	4.1	936	22	AAF58262	Oligonucleotide D2
20	171.2	4.1	936	22	AAF58255	Oligonucleotide D1
21	168	4.0	936	22	AAF58252	Oligonucleotide D1
22	168	4.0	936	22	AAF58254	Oligonucleotide D1
23	168	4.0	936	22	AAF58257	Oligonucleotide D1
24	168	4.0	936	22	AAF58259	Oligonucleotide D2
25	168	4.0	936	22	AAF58262	Oligonucleotide D2
26	168	4.0	936	22	AAF58255	Oligonucleotide D1
27	163.2	3.9	364	21	AAC03267	Human secreted pro
28	161.2	3.9	169	16	AAT26750	Human gene signatu
29	137.4	3.3	500	22	AAI42184	Probe #10870 used
30	136	3.3	136	22	AAI55269	Probe #23955 used
31	121.8	2.9	424	21	AAH31011	Human colon cancer
32	99.4	2.4	300	21	AAA01526	Human colon cancer
33	63.4	1.5	244	22	AAF58238	Oligonucleotide D1
34	63.4	1.5	244	22	AAF58238	Oligonucleotide D1
35	50.8	1.2	612	22	AAH71471	Human cervical can
36	50.6	1.2	4590	22	AAH24065	Yeast AOB9604-asso
37	47.2	1.1	318	18	AAT91312	Human J143-1 secre
38	47.2	1.1	318	18	AAV00436	3' fragment of clo
39	46	1.1	2657	22	AAH16026	Human CDNA sequenc
40	46	1.1	513445	22	AAI61373	Soybean 318013 reg
41	44.8	1.1	518	14	AAQ48590	Alcohol oxidase ge
42	44.8	1.1	14796	19	AAV27941	Survivin gene. Ho
43	44.6	1.1	598	22	AAO5439	Mammalian vestibul
44	44.2	1.1	335913	22	AAI61371	Soybean 240017 reg
45	44.2	1.1	335913	22	AAI61372	Soybean 240017 reg

ALIGNMENTS

RESULT 1

AAC65900	
ID	AAC65900 standard; cDNA; 4181 BP.
XX	
AC	AAC65900;
XX	
XX	
DT	21-FEB-2001 (first entry)
XX	
DE	Human lung cancer-associated cDNA L523S.
XX	
KW	Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW	vaccine; detection; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200061612-A2.
XX	
PD	19-OCT-2000.
XX	
PF	03-APR-2000; 2000WO-US08896.
XX	
PR	02-APR-1999; 99US-0285479.
PR	17-DEC-1999; 99US-0466396.
PR	30-DEC-1999; 99US-0476496.
PR	10-JAN-2000; 2000US-0480884.
PR	22-FEB-2000; 2000US-0510376.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Wang T, Fan L;
XX	
DR	WPI; 2000-628399/60.
DR	P-PSDB; AAB11328.
XX	

1801 aagtgacagagtgtgttgcctcctgtagaccagacacgtgatgaaatgaccaagtgtgtc 1860
 1861 caaaataaactggtcactctatgcttgccaggttgccagagaaaaattcagaaattct 1920
 1866 caaaataaactggtcactctatgcttgccaggttgccagagaaaaattcagaaattct 1920
 1921 gactcaggtaagcagcaccacaacagagaggtctctgcaaaagtgcaccacctcagtcag 1980
 1921 gactcaggtaagcagcaccacaacagagaggtctctgcaaaagtgcaccacctcagtcag 1980
 1981 acggaagttaagcgtcaggaacagccacacagagcagagtgccaaacccaagaacaga 2040
 1981 acggaagttaagcgtcaggaacagccacacagagcagagtgccaaacccaagaacaga 2040
 2041 ttgcttaacccaacagatggcgctgacccctctacaggaatcacatgacacaagtgtttac 2100
 2041 ttgcttaacccaacagatggcgctgacccctctacaggaatcacatgacacaagtgtttac 2100
 2101 ctagccagttgttctgaggaacaggcaacttttgaactcctgtctctgtggaatgtat 2160
 2101 ctagccagttgttctgaggaacaggcaacttttgaactcctgtctctgtggaatgtat 2160
 2161 actttatgctctgaaatgtagacacccagcttttaaaacaaaacaaacaaacaaa 2220
 2161 actttatgctctgaaatgtagacacccagcttttaaaacaaaacaaacaaacaaa 2220
 2221 aaaaagggtgggaggggaaagagaagagctctgcacttccctttgttagtctca 2280
 2221 aaaaagggtgggaggggaaagagaagagctctgcacttccctttgttagtctca 2280
 2281 cagataacagatatcttaattctcttaattcccccataatgccagaaattggctta 2340
 2281 cagataacagatatcttaattctcttaattcccccataatgccagaaattggctta 2340
 2341 atgagtcttctactaaatcacaatagatgctccttaaatccaaattgttaaaattgga 2400
 2341 atgagtcttctactaaatcacaatagatgctccttaaatccaaattgttaaaattgga 2400
 2401 tcagataaatatcacaggaacttaaatgttaagccattagcatagaaaaactgttctca 2460
 2401 tcagataaatatcacaggaacttaaatgttaagccattagcatagaaaaactgttctca 2460
 2461 gttttatttttacctaaactaacatgagtaacctaaagggaagtgcgaattgttggc 2520
 2461 gttttatttttacctaaactaacatgagtaacctaaagggaagtgcgaattgttggc 2520
 2521 aggggtattaaacgtgcatttttactccaactcctcaggtatttcagtaatacaatgaaa 2580
 2521 aggggtattaaacgtgcatttttactccaactcctcaggtatttcagtaatacaatgaaa 2580
 2581 gcaaaattgttcccttttttggaaaatttatatacttttataatgatagaaagtcacaacg 2640
 2581 gcaaaattgttcccttttttggaaaatttatatacttttataatgatagaaagtcacaacg 2640
 2641 ttttttaaaaaataaatttaaaatttaacagcaatcagctaacaggcgaattaaagtattt 2700
 2641 ttttttaaaaaataaatttaaaatttaacagcaatcagctaacaggcgaattaaagtattt 2700
 2701 tactcttgctgggtgacagtaaaagctggaaaatttaatttcagggtttttttgaggttttg 2760
 2701 tactcttgctgggtgacagtaaaagctggaaaatttaatttcagggtttttttgaggttttg 2760
 2761 acacagttattagttaaatcaaatgttcaaaaatacaggagcagtgccctagatctggaga 2820
 2761 acacagttattagttaaatcaaatgttcaaaaatacaggagcagtgccctagatctggaga 2820
 2821 gcagcactaccatttattctttcatttatagttgggaaagtttttgacggtactacaaca 2880
 2821 gcagcactaccatttattctttcatttatagttgggaaagtttttgacggtactacaaca 2880
 2881 gtggtgcgaggagatttgggaacgctggtttaattggcttcagagagacttcagttttt 2940

Db 2881 gtggtgcgaggagattttggaacgctggtttaattggcttcaggagacttcagttttt 2940
 Qy 2941 gtttaagctacatgatgaatgataataaagcttcttctgcttctgactatcaatacctaa 3000
 Db 2941 gtttaagctacatgatgaatgataataaagcttcttctgcttctgactatcaatacctaa 3000
 Qy 3001 agaaaatgcacatcagtgaaagatgcaagactttcaactgactgggcaaaaagcaagcttta 3060
 Db 3001 agaaaatgcacatcagtgaaagatgcaagactttcaactgactgggcaaaaagcaagcttta 3060
 Qy 3061 gcttcttctataggatgcttagtttgcactacacatcagaccaatgggacagctataga 3120
 Db 3061 gcttcttctataggatgcttagtttgcactacacatcagaccaatgggacagctataga 3120
 Qy 3121 tgggtgacagtgctttaaacgcaaaaaggtcacatttccatggggccagcactgtcat 3180
 Db 3121 tgggtgacagtgctttaaacgcaaaaaggtcacatttccatggggccagcactgtcat 3180
 Qy 3181 gagctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaaaa 3240
 Db 3181 gagctcactaagctattttgaagatttttaagcactgataaattaaaaaataaaaaaaa 3240
 Qy 3241 aaattagactccacttaagtagtaaaagtataacagagatttctgtactgtgcaatcag 3300
 Db 3241 aaattagactccacttaagtagtaaaagtataacagagatttctgtactgtgcaatcag 3300
 Qy 3301 ttctttgaaaaaaaagctcaaaagatagagaatacaagaaaagttttngggataataattg 3360
 Db 3301 ttctttgaaaaaaaagctcaaaagatagagaatacaagaaaagttttngggataataattg 3360
 Qy 3361 aatgactgtgaaaaacatatgacctttgataaacgaactcatttgcctcactccttgacagca 3420
 Db 3361 aatgactgtgaaaaacatatgacctttgataaacgaactcatttgcctcactccttgacagca 3420
 Qy 3421 aagccagtcagtcacaaattgtgttgggttgggttgcctccaaggccagcgtctctctg 3480
 Db 3421 aagccagtcagtcacaaattgtgttgggttgggttgcctccaaggccagcgtctctctg 3480
 Qy 3481 aattgtatttttgagtttttggnttgaagatgatacagacnctgttacactgactctnaa 3540
 Db 3481 aattgtatttttgagtttttggnttgaagatgatacagacnctgttacactgactctnaa 3540
 Qy 3541 ggacatatnttataaccccttttaaaaaaaatccccctgctcattcttatttcgagatga 3600
 Db 3541 ggacatatnttataaccccttttaaaaaaaatccccctgctcattcttatttcgagatga 3600
 Qy 3601 atttcagacagactagatgtcttctgaaagatcaaatagacattntgaaaaatgatttaa 3660
 Db 3601 atttcagacagactagatgtcttctgaaagatcaaatagacattntgaaaaatgatttaa 3660
 Qy 3661 agtggttttcccttaagtctctgaaaaaagtttcttctgtagttttaacccaaaaaagtg 3720
 Db 3661 agtggttttcccttaagtctctgaaaaaagtttcttctgtagttttaacccaaaaaagtg 3720
 Qy 3721 cctttttgtcactgggttctctcactagcattcatgatttttttttcacacaaatgaattaa 3780
 Db 3721 cctttttgtcactgggttctctcactagcattcatgatttttttttcacacaaatgaattaa 3780
 Qy 3781 attgctaaaaatcaatgacatggcgttcttctggtgatttcaggtgaagtgtgtttaaggcca 3840
 Db 3781 attgctaaaaatcaatgacatggcgttcttctggtgatttcaggtgaagtgtgtttaaggcca 3840
 Qy 3841 gagctttctcagttatttgatttttttcccccaatttttgaatttttaaaaaatacacat 3900
 Db 3841 gagctttctcagttatttgatttttttcccccaatttttgaatttttaaaaaatacacat 3900
 Qy 3901 aggagctgcattttaaaccctgctggttttaaatctctgcanaatttcactcttagcctttta 3960
 Db 3901 aggagctgcattttaaaccctgctggttttaaatctctgcanaatttcactcttagcctttta 3960
 Qy 3961 gtatgcncaatcanaatttacttttactttaagcatttggagatttggagattctcggacta 4020
 Db 3961 gtatgcncaatcanaatttacttttactttaagcatttggagatttggagattctcggacta 4020

QY 4021 gctaagaaataattcnaataattgagttgttactcncncaanaatggttcattctctcatgn 4080
|||||
Db 4021 gctaagaaataattcnaataattgagttgttactcncncaanaatggttcattctctcatgn 4080
QY 4081 ataagtncncccccaatcagcttcatcttccaganaaccttgacgcagataaaatttttc 4140
|||||
Db 4081 ataagtncncccccaatcagcttcatcttccaganaaccttgacgcagataaaatttttc 4140
QY 4141 atcattaggtcccccacaaaaaaaccccccccccccccccccccccccccccccccc 4181
|||||
Db 4141 atcattaggtcccccacaaaaaaaccccccccccccccccccccccccccccccccc 4181

RESULT 2

AAZ36150

XX ID AAZ36150 standard; DNA; 4159 BP.

XX AC AAZ36150;

XX DT 11-FEB-2000 (first entry)

XX DE DNA encoding cancer associated antigen KOC-1.

XX KW Cancer associated antigen; KOC-1; cancer; vaccine; CT7; ss.

XX OS Homo sapiens.

XX PN W09954738-A1.

XX PD 28-OCT-1999.

XX PF 16-MAR-1999; 99WO-US05766.

XX PR 17-APR-1998; 98US-0061709.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX DR WPI; 2000-013284/01.

XX PT Nucleotides representing cancer-associated genes, used to develop
XX products for the diagnosis, monitoring and treatment of cancers -

XX PS Claim 88; Page 39-40; 44pp; English.

XX CC The present sequence represents a cancer associated antigen gene
CC designated KOC-1. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polypeptide was isolated from
CC SK-MEL-37 melanoma cells. The polypeptide has some homology with
CC MAGE-10, limited to about 210 carboxy terminal amino acids. The amino
CC terminal of the protein has a repetitive pattern, with repeats rich in
CC serine, proline, glutamine and leucine, and an almost invariable core of
CC the peptide given in AAY43877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
CC stimulating factor (GM-CSF).

XX SQ Sequence 4159 BP; 1281 A; 830 C; 851 G; 1181 T; 16 other;

Query Match 99.1%; Score 4143; DB 21; Length 4159;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 4159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtggatgcgttgggttagctagctttttctttctttctttcttttaaacacatct 60
|||||
Db 1 ggtggatgcgttgggttagctagctttttctttctttctttcttttaaacacatct 60

QY 61 agacaagaaaaaacagcctcgatctgatttttcaactcctcgtcttcttgcttggttc 120
Db 61 agacaagaaaaaacagcctcgatctgatttttcaactcctcgtcttcttgcttggttc 120
QY 121 ttactgtttgtgtattttaaagcgagaagcgaggggaacaaaacacagctgagatcca 180
Db 121 ttactgtttgtgtattttaaagcgagaagcgaggggaacaaaacacagctgagatcca 180
QY 181 tccatcacctgggtgtgttttaatttttctgtttttctcgttatttttttaacaacc 240
Db 181 tccatcacctgggtgtgttttaatttttctgtttttctcgttatttttttaacaacc 240
QY 241 actcttcaaatgaacaaaactgtatatcggaaacctcagcgagacgcgcgccctcgga 300
Db 241 actcttcaaatgaacaaaactgtatatcggaaacctcagcgagacgcgcgccctcgga 300
QY 301 cctagaagaattcttcaagacccaagatcccgtgtcgggaccccttctcgttgagagac 360
Db 301 cctagaagaattcttcaagacccaagatcccgtgtcgggaccccttctcgttgagagac 360
QY 361 tggctacgcgttcgtgactgcgcggagcgtgggcccctcaaggccatcagaggcgt 420
Db 361 tggctacgcgttcgtgactgcgcggagcgtgggcccctcaaggccatcagaggcgt 420
QY 421 ttcaggtaaaatagaactgcacgggaaacccatagaagttagcactcggctcccaaaaag 480
Db 421 ttcaggtaaaatagaactgcacgggaaacccatagaagttagcactcggctcccaaaaag 480
QY 481 gcaaggaattcggaaacttcagatacgaataatccccctcatttacagtgagggtgct 540
Db 481 gcaaggaattcggaaacttcagatacgaataatccccctcatttacagtgagggtgct 540
QY 541 ggaatgttactagtcagatggagtggtagagcgtgtgagcaagtgaacactgactc 600
Db 541 ggaatgttactagtcagatggagtggtagagcgtgtgagcaagtgaacactgactc 600
QY 601 ggaactgcagtgtaaatgaactatccagtaaggaaacccaagctagacaagcactaga 660
Db 601 ggaactgcagtgtaaatgaactatccagtaaggaaacccaagctagacaagcactaga 660
QY 661 caaactgaatggatttcagttagagaatttcaccttgaagctagcctatccctgatga 720
Db 661 caaactgaatggatttcagttagagaatttcaccttgaagctagcctatccctgatga 720
QY 721 aatggccgccagcaaaccccttgccagcagcccgaggctgcgcggggttgggcagag 780
Db 721 aatggccgccagcaaaccccttgccagcagcccgaggctgcgcggggttgggcagag 780
QY 781 gggctcctcaaggcaggggtctccagatccgtatccaagcagaaacacatgtgattgcc 840
Db 781 gggctcctcaaggcaggggtctccagatccgtatccaagcagaaacacatgtgattgcc 840
QY 841 tctgcgcctcgtggttcccccacaaatttgttgagccatcataggaagaaagtgccac 900
Db 841 tctgcgcctcgtggttcccccacaaatttgttgagccatcataggaagaaagtgccac 900
QY 901 cattcggaacatcaccaaaaacagaccagctctaaatcgatgtccaccgtaaagaaatgc 960
Db 901 cattcggaacatcaccaaaaacagaccagctctaaatcgatgtccaccgtaaagaaatgc 960
QY 961 gggggctgctgagaagtcgatctactatcctctctactcctgaagcaccctctcggttg 1020
Db 961 gggggctgctgagaagtcgatctactatcctctctactcctgaagcaccctctcggttg 1020
QY 1021 taagtctattctggagattatgcataaggaagctcaagatatataatccacagaagat 1080
Db 1021 taagtctattctggagattatgcataaggaagctcaagatatataatccacagaagat 1080
QY 1081 cccctggaagatttagctcataataactttgttggcgctcttatttggtaaagaagaag 1140
Db 1081 cccctggaagatttagctcataataactttgttggcgctcttatttggtaaagaagaag 1140
QY 1141 aatcttataaaaaaattgagcaagacacagacactaaataatcacgatatctccttcgagga 1200

Db 1141 aatcttaaaaaattgagacagacagacactaaaatacagatatctccattgcagga 1200
Qy 1201 attgacgtgtataatccagaaacgacactattacaggttaaaggcaatgttgagacatgtgc 1260
Db 1201 attgacgtgtataatccagaaacgacactattacaggttaaaggcaatgttgagacatgtgc 1260
Qy 1261 caaagctgaggagatcatgaagaaaaatcaggagctcttatgaaaaatgatattgcttc 1320
Db 1261 caaagctgaggagatcatgaagaaaaatcaggagctcttatgaaaaatgatattgcttc 1320
Qy 1321 tatgaattccaagcacatttaattccctggattaaatcagccctcagccatgacctccoccta 1380
Db 1321 tatgaattccaagcacatttaattccctggattaaatcagccctcagccatgacctccoccta 1380
Qy 1381 acccaactcaggagtgccacctcccacactcagggcccccctcagccatgacctccoccta 1440
Db 1381 acccaactcaggagtgccacctcccacactcagggcccccctcagccatgacctccoccta 1440
Qy 1441 cccgcagtttgagcaatcagaacaggagactgttcatcagtttatcccagctctatcagt 1500
Db 1441 cccgcagtttgagcaatcagaacaggagactgttcatcagtttatcccagctctatcagt 1500
Qy 1501 cagtgccatcatcggcaagcaggccagcacatcaagcagctttctcgtttgtgagagc 1560
Db 1501 cagtgccatcatcggcaagcaggccagcacatcaagcagctttctcgtttgtgagagc 1560
Qy 1561 tscaatataagattgctccagcggaagcaccagatgctaaagtggagatgggtgattatcac 1620
Db 1561 tscaatataagattgctccagcggaagcaccagatgctaaagtggagatgggtgattatcac 1620
Qy 1621 tggaccaccagaggtcagttcaaggctcagggaagaatttatgaaaaaattaaagaaga 1680
Db 1621 tggaccaccagaggtcagttcaaggctcagggaagaatttatgaaaaaattaaagaaga 1680
Qy 1681 aaactttgttagcttaagaagaaggtgaaactbgaagctcatatcagagtgccatcctt 1740
Db 1681 aaactttgttagcttaagaagaaggtgaaactbgaagctcatatcagagtgccatcctt 1740
Qy 1741 tgcctgtgcagaggttatgtgaaaaggaggcaaaacggtgaatgaactcagaatttgc 1800
Db 1741 tgcctgtgcagaggttatgtgaaaaggaggcaaaacggtgaatgaactcagaatttgc 1800
Qy 1801 aagtcagaagttgttgccttcgtgacacagacactgatgagaatgaccaagtgttgt 1860
Db 1801 aagtcagaagttgttgccttcgtgacacagacactgatgagaatgaccaagtgttgt 1860
Qy 1861 caaataactggtcactctetagtctgcaggtgtcccagagaaaaatccaggaaattct 1920
Db 1861 caaataactggtcactctetagtctgcaggtgtcccagagaaaaatccaggaaattct 1920
Qy 1921 gactcaggtlaaagcagaccacaacagaggctctgcaaaagtggacacotcagtcaag 1980
Db 1921 gactcaggtlaaagcagaccacaacagaggctctgcaaaagtggacacotcagtcaag 1980
Qy 1981 acggaagtaaggtcaggaaacagccacacagagcgagatgccaaacaaagacaga 2040
Db 1981 acggaagtaaggtcaggaaacagccacacagagcgagatgccaaacaaagacaga 2040
Qy 2041 ttgcttaaccaacagatgggctgacccccctatccagaatccatgacaaagtttttac 2100
Db 2041 ttgcttaaccaacagatgggctgacccccctatccagaatccatgacaaagtttttac 2100
Qy 2101 ctagcaggttcttgaggaccaggaacacttttgaactctctgtctgtgagaatgtat 2160
Db 2101 ctagcaggttcttgaggaccaggaacacttttgaactctctgtctgtgagaatgtat 2160
Qy 2161 actttatgctctgaaatgtatgaccccagctttaaacacaaacaaacaaacaaa 2220
Db 2161 actttatgctctgaaatgtatgaccccagctttaaacacaaacaaacaaacaaa 2220
Qy 2221 aaaaaggttggggagggaagagagagctctgcacttccctttgtgtagtctca 2280
Db 2221 aaaaaggttggggagggaagagagagctctgcacttccctttgtgtagtctca 2280

Db 2221 aaaaaggttggggagggaagagagagctctgcacttccctttgtgtagtctca 2280
Qy 2281 cagtaaacagatatcttaattctcttaattatcccccataatgcagaaaattggctta 2340
Db 2281 cagtaaacagatatcttaattctcttaattatcccccataatgcagaaaattggctta 2340
Qy 2341 atgatgtcttcaactaaattcatcaaatagattgctcctaaatcccaattgttaaaaattgga 2400
Db 2341 atgatgtcttcaactaaattcatcaaatagattgctcctaaatcccaattgttaaaaattgga 2400
Qy 2401 tcagaataattatcacaggaaacttaaatgttaagccattagcatagaaaaactgttctca 2460
Db 2401 tcagaataattatcacaggaaacttaaatgttaagccattagcatagaaaaactgttctca 2460
Qy 2461 gttttatttttaactaaactaacatgagtaacctaaagggaagtgctgaatggtgtgc 2520
Db 2461 gttttatttttaactaaactaacatgagtaacctaaagggaagtgctgaatggtgtgc 2520
Qy 2521 aggggtattaaacgtgcatttttactcaactcctcaggtattcagtaatacaatgaaa 2580
Db 2521 aggggtattaaacgtgcatttttactcaactcctcaggtattcagtaatacaatgaaa 2580
Qy 2581 gaaaaattgttctctttttttgaaaaatttataactttataatgatagaaagccaacgg 2640
Db 2581 gaaaaattgttctctttttttgaaaaatttataactttataatgatagaaagccaacgg 2640
Qy 2641 ttttttaaaaaataaatttaaaatttaacagcaatcaacagcaatacaggaataaagatttt 2700
Db 2641 ttttttaaaaaataaatttaaaatttaacagcaatacagcaatacaggaataaagatttt 2700
Qy 2701 tacttctggctggtagacagtaaaagctggaaaatttaatttcagggttttttgaggctttg 2760
Db 2701 tacttctggctggtagacagtaaaagctggaaaatttaatttcagggttttttgaggctttg 2760
Qy 2761 acacagttattagttaaatcacaatgttcaaaaatacaggagcagtgctagatctggaga 2820
Db 2761 acacagttattagttaaatcacaatgttcaaaaatacaggagcagtgctagatctggaga 2820
Qy 2821 gcagcaactaccatttattcttccattatagttgggaagtttttgacggttaactaaca 2880
Db 2821 gcagcaactaccatttattcttccattatagttgggaagtttttgacggttaactaaca 2880
Qy 2881 gtggtcgcaggagattttggaaacgctgtttaaatggcttcaggagacttcagtttttt 2940
Db 2881 gtggtcgcaggagattttggaaacgctgtttaaatggcttcaggagacttcagtttttt 2940
Qy 2941 gtttggtacatgattgaatgcataataaattgctttgtctctgactatcaatacctaa 3000
Db 2941 gtttggtacatgattgaatgcataataaattgctttgtctctgactatcaatacctaa 3000
Qy 3001 aqaaaagtcacagtgaaagatgcaagacttccaactgctggaagacttcaactgctggaagacttca 3060
Db 3001 aqaaaagtcacagtgaaagatgcaagacttccaactgctggaagacttcaactgctggaagacttca 3060
Qy 3061 gcttcttcttaggtgcttagttggccactacacttcagaccatgggacagtcataga 3120
Db 3061 gcttcttcttaggtgcttagttggccactacacttcagaccatgggacagtcataga 3120
Qy 3121 tgggtgacagtgattttaaacycaaaaaggctacatttccattggggccagcactgcat 3180
Db 3121 tgggtgacagtgattttaaacycaaaaaggctacatttccattggggccagcactgcat 3180
Qy 3181 gggcctcactaagctattttgaagatttttaagcactgataaataaaaaaataaaaaa 3240
Db 3181 gggcctcactaagctattttgaagatttttaagcactgataaataaaaaaataaaaaa 3240
Qy 3241 aaattagactccacttaagtagtaaaagtataacagatttctgtatactgtgcaatcag 3300
Db 3241 aaattagactccacttaagtagtaaaagtataacagatttctgtatactgtgcaatcag 3300
Qy 3301 ttctttgaaaaaaagtcaaaaagatagagaatacaagaaaagtttngggatataatttg 3360
Db 3301 ttctttgaaaaaaagtcaaaaagatagagaatacaagaaaagtttngggatataatttg 3360

QY 3361 aatgactgtgaaacacatataacacattgacaccttctgataacgaactcatttgcctcactccttgacagca 3420
Db 3361 aatgactgtgaaacacatataacacattgacaccttctgataacgaactcatttgcctcactccttgacagca 3420
QY 3421 aagccagtagctacataattgttgggtgtgggtgtgtcctcaaggccacgctgctctctg 3480
Db 3421 aagccagtagctacataattgttgggtgtgggtgtgtcctcaaggccacgctgctctctg 3480
QY 3481 aattgattttttgagtttggnttgnnaagatgatacagacnactgttacactgatctctnaa 3540
Db 3481 aattgattttttgagtttggnttgnnaagatgatacagacnactgttacactgatctctnaa 3540
QY 3541 ggacataatntataaacacctttataaaaaaaatccctgcctcattcttatttgcgagatga 3600
Db 3541 ggacataatntataaacacctttataaaaaaaatccctgcctcattcttatttgcgagatga 3600
QY 3601 attcgcatacagactagatgtcttctgaagatcaatagacattntgaaatgatttaa 3660
Db 3601 attcgcatacagactagatgtcttctgaagatcaatagacattntgaaatgatttaa 3660
QY 3661 agtgtttctctaatgttctctgaaacaaagtcttctttgttagttttaacccaaaaaagtg 3720
Db 3661 agtgtttctctaatgttctctgaaacaaagtcttctttgttagttttaacccaaaaaagtg 3720
QY 3721 cccctttgtcactggtttctcctagcattcatgatttttttttccacacaaatgaattaa 3780
Db 3721 cccctttgtcactggtttctcctagcattcatgatttttttttccacacaaatgaattaa 3780
QY 3781 attgctaaatcatcagactggtcttctggtgatttcaggtaagatgtttaaaggcca 3840
Db 3781 attgctaaatcatcagactggtcttctggtgatttcaggtaagatgtttaaaggcca 3840
QY 3841 gagctttctcagattgtgatttttttccccaatattgtgatttttttaaaaaatacacat 3900
Db 3841 gagctttctcagattgtgatttttttccccaatattgtgatttttttaaaaaatacacat 3900
QY 3901 agagctgcatttataaaccttgcgtgttttaattctgtcnaattcactctagcctttta 3960
Db 3901 agagctgcatttataaaccttgcgtgttttaattctgtcnaattcactctagcctttta 3960
QY 3961 gtatggcnaatcanaatttacttttacttaagcatttgttaatttgagtagtctgacta 4020
Db 3961 gtatggcnaatcanaatttacttttacttaagcatttgttaatttgagtagtctgacta 4020
QY 4021 gctaagaaataattcnaataattgattttgtactcncnaaanatgggtcattctcatgn 4080
Db 4021 gctaagaaataattcnaataattgattttgtactcncnaaanatgggtcattctcatgn 4080
QY 4081 ataattgtnccccaatgcagcttcaattttccaganaaccttgacgcaggataaatttttc 4140
Db 4081 ataattgtnccccaatgcagcttcaattttccaganaaccttgacgcaggataaatttttc 4140
QY 4141 atcatttaggtccccaaaa 4159
Db 4141 atcatttaggtccccaaaa 4159

RESULT 3

AAC66035

ID AAC66035 standard; cDNA; 1740 BP.

XX AAC66035;

AC AAC66035;

XX AAC66035;

DT 21-FEB-2001 (first entry)

XX Human lung cancer-associated cDNA antigen L523S.

XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;

KW vaccine; detection; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

PN WO200061612-A2.

XX 19-OCT-2000.

XX 03-APR-2000; 2000WO-US08896.

XX 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX (CORI-) CORIXA CORP.

PA Wang T, Fan L;

XX WPI: 2000-628399/60.

XX P-PSDB; AAB11365.

DR Isolated polypeptide comprising an immunogenic portion of a lung tumor

DR protein is used for detecting and monitoring progression of lung cancer

XX in a patient

XX Claim 1a; Page 258-259; 261pp; English.

PS This invention describes a novel isolated polypeptide (I) which

XX comprising an immunogenic portion of a lung tumor protein or variant (P2)

CC which have cytostatic activity. The polypeptides and polynucleotides are

CC used in compositions and vaccines to inhibit the development of cancer,

CC especially lung cancer, in a patient. Methods described in the invention

CC can be used to monitor the progression of a cancer by carrying out the

CC detection at subsequent time points and comparing the results from the

CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient

CC are treated with P2, polynucleotides encoding P2 or antigen presenting

CC cells expressing P2 and then administered to the patient to inhibit

CC development of cancer.

XX Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 other;

Query Match 41.5%; Score 1736.8; DB 21; Length 1740;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1738; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 251 atgaacaaactgtatatcggaacacctcagcgagaacgcccccctcgagacctagaaagt 310

Db 1 atgaacaaactgtatatcggaacacctcagcgagaacgcccccctcgagacctagaaagt 60

QY 311 atcttcaagagacccaagatcccgggtgtcgggaccttctctgtgaagactggctacgcy 370

Db 61 atcttcaagagacccaagatcccgggtgtcgggaccttctctgtgaagactggctacgcy 120

QY 371 ttcgtggactgcccgacgagagactggccctcaagccatcgagcgcttccaggttaa 430

Db 121 ttcgtggactgcccgacgagagactggccctcaagccatcgagcgcttccaggttaa 180

QY 431 atagaactgcacgggaaacccatagaaagttgagcactcgtcccaaaaaggcaaggatt 490

Db 181 atagaactgcacgggaaacccatagaaagttgagcactcgtcccaaaaaggcaaggatt 240

QY 491 cggaaacttcagatacagaaatatcccgcctcatttccagtgaggaggtgctggatagttta 550

Db 241 cggaaacttcagatacagaaatatcccgcctcatttccagtgaggaggtgctggatagttta 300

QY 551 ctagtccagtatggagtgtggagagctgtgagcaagtgaacactgactcggaaactgca 610

Db 301 ctagtccagtatggagtgtggagagctgtgagcaagtgaacactgactcggaaactgca 360

QY 611 gttgtaaatgtaacctatttccagtaaggaccagctagacaaagcactagacaaactgaat 670

Db 361 gttgtaaatgtaacctatttccagtaaggaccagctagacaaagcactagacaaactgaat 420

QY 671 ggaatttcagttagagaatttcaaccttgaaagtgcctatatccctgatgaaatggccgcc 730

Db 421 gattttcagtttagaagaatttcaccttgaaagtacgtctataccctgtagaaacggcgccc 480
Qy 731 cagcaaaaaccccttcagcagcccgaggtgcgcggggcttgagcagaggggctccctca 790
Db 481 cagcaaaaaccccttcagcagcccgaggtgcgcggggcttgagcagaggggctccctca 540
Qy 791 aggcagggttcacaggtacggtatcccaagcagaaacacatgtgattgcctctgcgctg 850
Db 541 aggcagggttcacaggtacggtatcccaagcagaaacacatgtgattgcctctgcgctg 600
Qy 851 ctgggtcccccacccattgttgagcccatcatagaaagaagggtgcacccattcggaac 910
Db 601 ctgggtcccccacccattgttgagcccatcatagaaagaagggtgcacccattcggaac 660
Qy 911 atcaccaaacagaccagctctaaaatcgatgtccaccgttaaaagaataatgcggggctgct 970
Db 661 atcaccaaacagaccagctctaaaatcgatgtccaccgttaaaagaataatgcggggctgct 720
Qy 971 gagaagtcgattactatctctactctctactctgaaggcacctctgcgctgtgaagtctatt 1030
Db 721 gagaagtcgattactatctctactctctactctgaaggcacctctgcgctgtgaagtctatt 780
Qy 1031 ctggagattatgcataaagaagctcaagataataataattcacagaagagatcccccttgaag 1090
Db 781 ctggagattatgcataaagaagctcaagataataataattcacagaagagatcccccttgaag 840
Qy 1091 attttagctcataataaactttgttgagcgtcttattgttaagaagaagaatactttaa 1150
Db 841 attttagctcataataaactttgttgagcgtcttattgttaagaagaagaatactttaa 900
Qy 1151 aaaattgagcagacacagacactaaaatcacgatatctccattgcaggaattgcgctg 1210
Db 901 aaaattgagcagacacagacactaaaatcacgatatctccattgcaggaattgcgctg 960
Qy 1211 tataatccagaacgcactattacagtttaaaaggcaatgttgagacatgtgcacaaagctgag 1270
Db 961 tataatccagaacgcactattacagtttaaaaggcaatgttgagacatgtgcacaaagctgag 1020
Qy 1271 gagagatcatgaagaaatacaggagctcttatgaaataatgtattgtctctatgaattct 1330
Db 1021 gagggagatcaagaagaaatacaggagctcttatgaaataatgtattgtctctatgaattct 1080
Qy 1331 caagcacatttaattccctggattaaatctgaacgccttgggtctgttcccccacttca 1390
Db 1081 caagcacatttaattccctggattaaatctgaacgccttgggtctgttcccccacttca 1140
Qy 1391 gggatgcacactccacactcagggcccccttcagccatgactctctccctacccgcagttt 1450
Db 1141 gggatgcacactccacactcagggcccccttcagccatgactctctccctacccgcagttt 1200
Qy 1451 gagcaatcagaaacggagactgttcacagtttattccagctctctacagtcgggtccatc 1510
Db 1201 gagcaatcagaaacggagactgttcacagtttattccagctctctacagtcgggtccatc 1260
Qy 1511 atcggaagcaggggcagcacatcacagcgtttctctgttctgtggagcttcaattaaag 1570
Db 1261 atcggaagcaggggcagcacatcacagcgtttctctgttctgtggagcttcaattaaag 1320
Qy 1571 attgtccagcgggaagcaccagatgctaaagtgaaggatggtgattatcactggaccacca 1630
Db 1321 attgtccagcgggaagcaccagatgctaaagtgaaggatggtgattatcactggaccacca 1380
Qy 1631 gaggtcagttcaaggtcagggagaatttatgaaataataaagaagaatactttgtt 1690
Db 1381 gaggtcagttcaaggtcagggagaatttatgaaataataaagaagaatactttgtt 1440
Qy 1691 agtctaaagaagaggtgaaactgaaactcatatcacagtgccatctttgtgctgctgc 1750
Db 1441 agtctaaagaagaggtgaaactgaaactcatatcacagtgccatctttgtgctgctgc 1500
Qy 1751 agagtatttgaaagggggcaaacggtgaatgaacttcagaaattttgtcaagtcagaa 1810

Db 1501 eqagttatttgaaaaaggagggaacacggtgaatgaacttcagaatttgcgaagtcagaa 1560
Qy 1811 gttgtgtcccttcgtagccagacacactgatgagaatgaccagtggtgtgtcacaataact 1870
Db 1561 gttgtgtcccttcgtagccagacacactgatgagaatgaccagtggtgtgtcacaataact 1620
Qy 1871 ggtcactcttatgcttgcaggttcccagagagaaaaattcaggaaattctgactcaggtta 1930
Db 1621 ggtcactcttatgcttgcaggttcccagagagaaaaattcaggaaattctgactcaggtta 1680
Qy 1931 aagcagcaccacaacagaagctctcaaaagtgaaccacctcagtcagacggaagtaa 1990
Db 1681 aagcagcaccacaacagaagctctcaaaagtgaaccacctcagtcagacggaagtaa 1740
RESULT 4
AAZ10617
ID AAZ10617 standard; cDNA; 2224 BP.
XX
AC AAZ10617;
XX
DT 17-NOV-1999 (first entry)
XX
DE cDNA encoding a murine c-myc coding region determinant binding protein.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor;
KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 131..1864
FT /*tag= a
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99MO-US04897.
XX
PR 09-MAR-1998; 98US-0077372.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
DR P-PSDB; AAY30649.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
PT coding region determinant-binding protein -
PS Example; Fig 1A-D; 79pp; English.
XX
CC The present sequence encodes a murine c-myc coding region determinant
CC binding protein (CRD-BP). The presence or absence of a tumor can be
CC determined by determining the levels of CRD-BP present in the suspect
CC tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
CC and so prolongs its half-life. The methods are used for diagnosing
CC presence or absence of a tumor in a human, especially breast, colon
CC and pancreatic cancer. They are also used to inhibit cancer cell
CC growth.
XX
SQ Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 other;

Query Match 19.2%; Score 804.4; DB 20; Length 2224;
Best Local Similarity 67.0%; Pred. No. 1.2e-170;
Matches 1141; Conservative 0; Mismatches 561; Indels 0; Gaps 0;
Qy 247 cacaatgaacaaactgtatatcggaaacctcagcgagaacgcgcgccctcggaactaga 306
||| ||||||| || || ||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 127 caccatgaacaagctttacatcgtggcgaacctcaacgagagtgtagcccccgagacttggga 186
Qy 307 aagtattctcaagcagcccaagatccccggtgtcgggacccttctcgttgaaagactgcta 366
Db 187 gaaagtattcgcggagcacaagaatctctcagcggccagttcttgggtcaaatccgcta 246
Qy 367 cgccttcgtgactgcgcggagcagagactgggcccctcaagccatcagcgcgttttcagg 426
Db 247 cgccttcgtggattgcgccacagcactggcgatgaaggccatcgaaaactttctcggg 306
Qy 427 taaaatagaactgcacgggaaacccatagaagttgagcactcgttccccaaaaagggcaag 486
Db 307 gaaagtgaactgcaaggaaaaacgtctagagattgaacactcagtcctcccaaaaaacaag 366
Qy 487 gattcggaaacttcagatagaataatccccgcctcatttacagtgggagtgctgggatag 546
Db 367 gagtgcgaaaaatcacagatccgcgaatatctccacctcagtcctcgatgggaagtgcctagatag 426
Qy 547 ttactagtcagatagagtggtggagagctgtgagcaagtgaacactgactcgggaaac 606
Db 427 cctgctggctcagtcacgttacagtgggaactgtgaaagtgaacactgaaagtgaac 486
Qy 607 tgcagttgtaaatgaacctattccagtaagaccagagctagacaagcactagacaact 666
Db 487 agcgtgtcaactcaactactactaaccgaggagcagaccaggcaagctatcatgaagct 546
Qy 667 gaatggatttcagtagagaatttcaaccttgaagtagtcctatatcctcgtgatgaattggc 726
Db 547 aaatggccataactggagaacctgccttgaaggtctctcatatcctatgagcagat 606
Qy 727 cggccagcaaaaccttgcagcagccccgaggtcgccgggggcttggcagaggggctc 786
Db 607 aacacaggttcctgagaaatggggctgtgaggctttgggtcgtggggccagccccggca 666
Qy 787 ctcaaggcagggttctccagagctcgtatccaagcagaacaccatgtgatttgccttcgcg 846
Db 667 aggttcgccgtggcagcaggggtccagccaagcagcagccagtgacatccctctccg 726
Qy 847 cctgctggttcccaaccaatttgttgagccatcatataggaaaagaagtgccaccattcg 906
Db 727 gctcctggtgctacgcagtatgtaggcgtatcatcttgcgaaggaggtgccaccatccg 786
Qy 907 gaacatcaccaaacagaccagctcaaaatcgatgtccaccgttaaaagaataatcgggggc 966
Db 787 aaacatcacaaacagcagctccaaataatagacgtgcatagaaaggagaatcggggcgc 846
Qy 967 tgcgtgaagtcgattactactcctctactcctcgaaggccacctcgtcggcttgtgaagtc 1026
Db 847 tgcgagaagccatcagctgcatccaacctgaaggctcctcctccgcgtgcaagat 906
Qy 1027 tattctggagattatgcataagaagctcaagatataataattcacagaagatccccctt 1086
Db 907 gatctggagattatgcacaaggaggcaaggacacacacacacggcagatgaagttccct 966
Qy 1087 gaagattttagctcaataaacttgttgacgtcttatttggtaaaagaagaagaatact 1146
Db 967 gaagatcctgctcataacaaactcgtcgggcgactcatttggcaaggagccggaaact 1026
Qy 1147 taaaaaattgagcaagcacacacataaaatcacgatatctcctcattgcagaaattgac 1206
Db 1027 gaagaaggtggagcagcagacagagcagaagatcaccatctcatcgtccagaccctac 1086
Qy 1207 gctgtataatccgaacgcactattacagttaaaggcaatgttgagacatgtgccaaagc 1266
Db 1087 gctctataacctgagaggaccatcacttgaaggcgccatigagaactgttgcagggc 1146
Qy 1267 tgaggaggagatcatgaagaaaatcaggaggtcttatgaaaaatgatattgtctctatgaa 1326
Db 1147 cgagcaggagatcatgaagaaagtcgagaggttaccagaacgacgtggccgcctag 1206
Qy 1327 tcttcagcacatttaattcctgagtaataatctgaacgcttgggtctgtctccacccac 1386
Db 1207 cttgcggtcccaactcctcctgggttaaccttgcgtgaggtctctctccagcttc 1266

RESULT 5

AAZ36152

ID AAZ36152 standard; DNA; 3412 BP.

XX

AC AAZ36152;

XX

DT 11-FEB-2000 (first entry)

XX

DE DNA encoding cancer associated antigen KOC-3.

XX

KW Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.

XX

OS Homo sapiens.

XX

PN WO9954738-A1.

XX

PD 28-OCT-1999.

XX

PF 16-MAR-1999; 99WO-US05766.

XX

PR 17-APR-1998; 98US-0061709.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX

XX WPI; 2000-013284/01.

PT Nucleotides representing cancer-associated genes, used to develop

PT products for the diagnosis, monitoring and treatment of cancers -

XX

Db	1057	gctcctttagcaggctcccgagcaggagatggtgcaggtgttttatccocgcccaggcag	11116
Qy	1500	tcggtgcacatcatcggaagcaggccagcacatacaagcagctttctcgcttgcgtggag	1559
Db	1117	tggcgccatcatcggaagaggcgagcacatacaacagctctccggtttgcacgcg	11776
Qy	1560	cttcaattaaagattgtctcaagcggaagcaccagatgctaaagtgaggatgagttatca	1619
Db	1177	ctctcataagattgcacacccgaacacctgactccaaagttcgatggttatcatca	1236
Qy	1620	ctggaccaccagaggtctgattcaaggtctcagggaagaatttatggaaaaataaagaag	1679
Db	1237	ctggaccgccagagggcccaattcaaggtctcagggaagaatctatggcaaacccaaggag	1296
Qy	1680	aaaactttgtatgctctaaagaagagggtgaaacttgaagcttcatatcagagtgccatct	1739
Db	1297	agaactctttgtgtcccaaggaggagtgaaagctggagaccacacatacgtgtgccagcat	1356
Qy	1740	tgtgctctgcagagttatttgaaaaagagggcacaacggtgaatgaaactcagaatttgt	1799
Db	1357	cagcagctggccgggtcatttggcaaaaggttgaaaaaacggtgaaacgagttgcagaatttga	1416
Qy	1800	caagtcgagaagtgtgtgtctctgtgcaccagacacctgtatgaaatgacaagtgagt	1859
Db	1417	cggcagctgaggttggttagtaccacaagagacagaccctgtatgagaagcaccaggtcatcg	1476
Qy	1860	tcaaaataactggtctacttctatgcttgcagaggttggccagagaaaaaattcagagaaattc	1919
Db	1477	tgaaaaatcatcggaacattctatgcccagtcagatggcttcaacgggaagatccgagacatcc	1536
Qy	1920	tgactcaggttaagcagcaccacaacag	1948
Db	1537	tggcccaggttaagcagcagcatcagaag	1565

RESULT 8

RESOLI
AAZ36151

AAZ36151
ID AAZ36151 standard; DNA: 1708 bp.

XX
XX

AAZ36151;

[illegible]

DT 11-FEB-2000 (first entry)

XX

XXXXXX

DE DNA encoding cancer associ

XX

KW Cancer associated antigen;

XX

OS Homo sapiens.

[illegible]

PN WO9954738-A1.

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

PD 28-OCT-1999.
yy

XX
PF 16-MAR-1999: 0040-1150576

PF 16-MAR-1999;
YY 99WO-US05/6

17-APR-1998 08US-006170

XX
PR I/-APR-1996; 980S-006170

PA (LUDW-) LUDWIG INST CANCER

XX
XX
FA / TUDMGT INSI CANCER

Chen Y., Gure A., Tsang S.

terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AY43877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony stimulating factor (GM-CSF).

Sequence 1708 BP: 447 A; 469 C; 473 G; 314 T; 5 other;

Query Match	13.9%	Score 580.6;	DB 21;	Length 1708;
Best Local Similarity	66.1%	Pred. No. 1.7e-120;		
Matches 857; Conservative	0;	Mismatches 434;	Indels 6;	Gaps 1;

Qy	652	agcactagacaaactgaatgatttcctagttagagaatttcaccttgaagcagcctatat	711
Db	37	agccatcatgaagctgaatgccaccagttggagaaacatgcctcgaggtctcctacat	96
Qy	712	ccctgatgaatgcccgcacagcaaaaccccttgcacagcccgcaggtgcgcggggcct	771
Db	97	cccgcgtgagcagatagcacagggaccctgagaaatggcgccgagggcgttggctctcg	156
Qy	772	tggccagagggcctcctcaagcgaggggtctccaggatccggtatcccaagcagaaaccatg	831
Db	157	gggtcagcccgcagggctcaccttggcagcgggggcccgccaagcagcagaagt	216
Qy	832	tgatttgctctgcgcctgctggttccaccaccaatttgttggagccatcatagaaaaaga	891
Db	217	ggatccccctcggtccttgggtgccaccacagtatgtgggtgcctattattggcaagga	276
Qy	892	agggtgccaccatcggaaacatcacaaacagaccagctctaaaatcgatgtccacogtaa	951
Db	277	gggggccacatcgcgaaacatcacaaacagaccagctccaagatagacgtgcataggaa	336
Qy	952	agaaatgcggggcgtctggaagtcgattactatctctctactctctgaaggcacctc	1011
Db	337	ggagaacgagtgcgctgtaaaagccatcagtgtcactccaccctgagggtgctc	396
Qy	1012	tgcgcttgtaagctattctggagattatgcataaaggaagctccaagatataaattcac	1071
Db	397	ctcgcgttgtaagatgactctggagataatgcataaagggcctaaaggacacaaacgc	456
Qy	1072	agaagagatcccccttgaagatttagctcataataacttgttggagcgtcttatgtgtaa	1131
Db	457	tgcgaggttccccctgaagatccctggccataataacttgtaggcgtctcataggcaa	516
Qy	1132	agaaggagaataatttataaaaaattggcgaagacacacagacactaaaatcacgatatcc	1191
Db	517	ggagagcggaacctgagaaggttagagcaagatcccgagacaaaaatcacctactctc	576
Qy	1192	attcaggaattgacgtgtataatcagacgcactattacagtttaaaggcaatgtga	1251
Db	577	gttcaagaccttaccctttacaaccttgagagaccatcacgtgaaggggccatcga	636
Qy	1252	gacatgtgcaaaagctgagggagagatcatgaagaaatcagggagctcttatgaaaaatga	1311
Db	637	gaattgtgcagcgccgagcaggaaataatgaagaaagttcgggagggcctatgagaatga	696
Qy	1312	tatgtctctatgaattctcaagcacatttaattctctgattaaatctgaacgccttggg	1371
Db	697	tgtggctgcaatga-----gctctacccttgatcccttggcctgaacctgggtcgtgag	750
Qy	1372	tctgttccccaccacttcagggatgccacatccccactcaggggcccccttcagccatgac	1431
Db	751	tcctttcccgacttcacccagcgagtcctccgcgcgctcccgagcagcgttactggggctgc	810
Qy	1432	tcctccctaccgcgagttttgagcaatcagaaacggagactgttcatcagtttatccacgc	1491
Db	811	tcctcatagctctcttatgcaggtctcccgagcagagaatgtgtgcaggtgtttatccccgc	870


```

Db 158 gctcaagataataaattcac--aagagatcccccttgagagatttagctcctaataaacttt 215
Qy 1112 gttgacgtcttattggttaagaaggaagaatactttaaaaaattgagcaagacacagac 1171
Db 216 gttgacgtcttattggttaagaaggaagaatactttaaaaaattgagcaagacacagac 275
Qy 1172 actaaatcacgatatctccattgcaggaattgacgctgtataatccagaacgcaactatt 1231
Db 276 actaaatcacgatatctccattgcaggaattgacgctgtataatccagaacgcaactatt 335
Qy 1232 acagttaaagccaattgttgagacatg-tgccaaagctgaggagagatcatgaagaaaaat 1290
Db 336 acagttaaagccaattgttgagacatgttgccaaagctgaggagagatcatgaagaaaaat 395
Qy 1291 caggagctctatgaaatgatattgctctctatgaattctcaagcacat 1339
Db 396 caggagctctatgaaatgatattgctctctatgaattctcaagcacat 444

RESULT 12
AAI56378
ID AAI56378 standard; DNA; 282 BP.
XX
AC AAI56378;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #25064 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 25064; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 282 BP; 77 A; 79 C; 70 G; 56 T; 0 other;
```

Query Match 6.7%; Score 282; DB 22; Length 282;
Best Local Similarity 100.0%; Pred. No. 7.1e-54;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 652 agcactagacaaaactgaatggatttcagtttagagaaatttcaccttgaaaagtagcctatat 711
Db 1 agcactagacaaaactgaatggatttcagtttagagaaatttcaccttgaaaagtagcctatat 60
Qy 712 ccttgatgaaatggcgcgcccaagcaaaaaccccttgacagagcccgaggtcgcgcgggggt 771
Db 61 ccttgatgaaatggcgcgcccaagcaaaaaccccttgacagagcccgaggtcgcgcgggggt 120
Qy 772 tgggcagagggggtcctcctcaagcaggggtctccagagatccgtatccaagcagaaaaaccatg 831
Db 121 tgggcagagggggtcctcctcaagcaggggtctccagagatccgtatccaagcagaaaaaccatg 180
Qy 832 tgatttgctctgcgctgctggttccaccaccaatttggtagccatcatgagaaaaaga 891
Db 181 tgatttgctctgcgctgctggttccaccaccaatttggtagccatcatgagaaaaaga 240
Qy 892 aggtgcccaccattcggaacatcaccaaacacagacccagctctaa 933
Db 241 aggtgcccaccattcggaacatcaccaaacacagacccagctctaa 282

RESULT 13
AAI43398
ID AAI43398 standard; DNA; 588 BP.
XX
AC AAI43398;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #12084 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 12084; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 588 BP; 156 A; 139 C; 143 G; 150 T; 0 other;
```

Query Match 6.7%; Score 282; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 9.5e-54;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	652	agcactagacaaactgaatgatttcagtttagagaatttcacctgaaagtagcctatat	711
Db	21	agcactagacaaactgaatgatttcagtttagagaatttcacctgaaagtagcctatat	80
Qy	712	ccctgatgaaatggccgcccaagcaaaaaccccttcagcagcagcccgaggctgcgcggggcct	771
Db	81	ccctgatgaaatggccgcccaagcaaaaaccccttcagcagcagcccgaggctgcgcggggcct	140
Qy	772	tgggcagaggggctcctcaagcgagggtctccaggatccgtatcccaagcagaaacccatg	831
Db	141	tgggcagaggggctcctcaagcgagggtctccaggatccgtatcccaagcagaaacccatg	200
Qy	832	tgatttgctctgcgctgctgttcacccacccaatttcttgagccatcataggaagaaga	891
Db	201	tgatttgctctgcgctgctgttcacccacccaatttcttgagccatcataggaagaaga	260
Qy	892	aggtgcacacattcgggaacatcaccaaacagaccagtcctaa	933
Db	261	aggtgcacacattcgggaacatcaccaaacagaccagtcctaa	302
RESULT 14			
AAA02565			
ID	AAA02565 standard; cDNA; 710 BP.		
XX	AAA02565;		
XX	19-MAY-2000 (first entry)		
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:2556.		
KW	Human; colon cancer; tumour; diagnosis; gene expression product;		
KW	probe; detection; cancerous state; metastasis; identification;		
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;		
XX	oestrogen receptor-negative breast cancer; lung cancer; ss.		
OS	Homo sapiens.		
XX	WO9958675-A2.		
XX	18-NOV-1999.		
XX	13-MAY-1999; 99WO-US10602.		
XX	14-MAY-1998; 98US-0085426.		
PR	15-MAY-1998; 98US-0085537.		
PR	15-MAY-1998; 98US-0085696.		
PR	21-OCT-1998; 98US-0105234.		
PR	27-OCT-1998; 98US-0105877.		
XX	(CHIR) CHIRON CORP.		
PA	(HYSE-) HYSEQ INC.		
XX	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;		
PI	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;		
PI	Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;		
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;		
XX	WPI; 2000-126369/11.		
DR			

cancerous state of the cell from which the test sample was derived.
The polynucleotides sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). This is particularly for breast cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-negative breast cancer, lung cancer, and colon cancer.

CC Sequence 710 BP; 199 A; 174 C; 180 G; 152 T; 5 other;

Query Match 6.3%; Score 263.2; DB 21; Length 710;
Best Local Similarity 67.5%; Pred. No. 1.7e-49;
Matches 370; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 801 ctccaggatcgcgtatccaagcagaacaccatgattggcttctcgctcgctggtccc 860
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 105 cccccggggcaccttcacgccagcagcagattgattccccgctgcgatcgtgccc 164
Qy 861 cccaatttgttgaggccatcatagaaaagaagtgcaccattcggaaacataccaca 920
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 165 cccagtttgttgccatcgcgaaggaggcttgaccataaagacatacctaagc 224
Qy 921 agaccagcttaaatcgatgccaccgttaaagaaatgcggggctgcgagaagctga 980
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 225 agaccagctccgggtagatatccatagaaaagagaactcggagctgcagagaagcgtg 284
Qy 981 ttactatctctactcctggaagccacctctgcggctttaaagtctattctggagatta 1040
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 285 tcaccatccatgccaccacagagggacttctgaagcatgcgcgatgattctgaaatca 344
Qy 1041 tgcataaggaagctcagagataaaattccagaagagatcccttgaagatttagctc 1100
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 345 tgcgaagagcagatgagaccaaacctagccgaagagattcctctgaaatcttggcac 404
Qy 1101 ataataactctgttgacgtcttatgtgtaagagagagaatacttaaaaaattgagc 1160
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 405 acatggtctgtgtggaagacgtattggaaaagaagcagaaaatttgagaaaattgaa 464
Qy 1161 aagcacagacactaaatacsgatatctccattgcagggaattgcacgtgtataatccag 1220
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 465 atgaacacaggaccaagaataacaatctcatcttgcaggatttgagcatacaaccgg 524
Qy 1221 aacgcactattacggttaaaaggcaattgtgagacatgtgccaaagctgaggagagatca 1280
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 525 aaagaacatcacitgtgaaggcagctgaggcctgtgccagctgcgagatagagatta 584
Qy 1281 tgaagaaaatcaggagctttatgaaaatgatatgtcttatgaactctcaagcacatt 1340
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 585 tgaagaactcgtgaggcccttgaaaatgatatgtcgtggttacgtaagtcctcaa 644
Qy 1341 taattcct 1348
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 645 tgccttct 652

RESULT .15
AAF58252/C
ID AAF58252 standard; DNA; 936 BP.
XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX OS gene expression; ss.
XX OS Synthetic.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2002, 14:54:23 ; Search time 19840.1 Seconds
(without alignments)
2264.506 Million cell updates/sec

Title: US-09-685-696-175

Perfect score: 4181

Sequence: 1 ggtgatgcgttgggtgtg.....aaaaaaaaaaaaaaaaaaaa 4181

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1680.6	40.2	2202	AK011689	AK011689 Mus muscu
2	842	20.1	904	10	AU117945
3	802	19.2	876	10	AU142785
4	801.2	19.2	2780	12	AK013940
5	759.2	18.2	808	10	AU117427
6	723.6	17.3	745	10	AU134323
7	716.2	17.1	902	11	BE869656
8	648.4	15.5	937	11	BI090231
9	638.8	15.3	723	10	AU127673
10	634	15.2	751	10	AU139887
11	630.4	15.1	790	10	BE545535
12	629.6	15.1	816	10	AU139815

13	617.6	14.8	805	11	BG108548
14	617	14.8	769	11	BF608004
15	610.6	14.6	633	11	BG499373
16	606.4	14.5	1101	11	BG576431
17	601.2	14.4	1005	11	BG435192
18	588.2	14.1	940	11	BF034031
19	578.8	13.8	754	11	BF215050
20	573.8	13.7	643	10	AU134724
21	572.8	13.7	590	11	BG434352
22	570.5	13.6	662	11	BF185290
23	570.2	13.6	654	11	BF185289
24	567	13.6	715	11	BG069090
25	545.8	13.1	565	10	AI807368
26	536.8	12.8	975	11	BG115140
27	531.8	12.7	580	10	AI086619
28	530.4	12.7	640	10	AW958019
29	529	12.7	898	11	BF664818
30	528.6	12.6	575	10	AW044616
31	524.6	12.5	766	11	BG434285
32	518.8	12.4	567	10	AW628546
33	514	12.3	587	11	BG079106
34	509.6	12.2	639	11	W81035
35	506	12.1	1144	11	BG531972
36	505.6	12.1	554	11	D63197
37	500.6	12.0	547	10	AW043682
38	495	11.8	495	10	AI088047
39	495	11.8	645	10	AA163596
40	494.6	11.8	501	10	AI824109
41	493.2	11.8	573	11	BG082113
42	492.4	11.8	847	11	BF697640
43	489	11.7	490	10	AI424833
44	488	11.7	712	11	BG528016
45	487.2	11.7	907	11	BF697335

ALIGNMENTS

RESULT 1

AK011689 LOCUS
DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610036B18, full insert sequence.
ACCESSION AK011689
VERSION AK011689.1 GI:12847972
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

AK011689 2202 bp mRNA HTC 05-JUL-2001
Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610036B18, full insert sequence.

AK011689 1 GI:12847972
CAP trapper.
Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library

Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2202)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)

99279253
10349636
2 (bases 1 to 2202)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
20499374
11042159
3 (bases 1 to 2202)
Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

181 AGGCAAACGGTGAATGNACTTCAGAAATTTGTC AAGTGCAGAAATTTGTTGTCCTCGTGA 240

QY 1828 ccagacacctgatgagatgaccaaagtggtgtgtaaaataactggtcacttctatgttg 1887
|||||
Db 241 CCAGACACCTGATGAGAAATGACCAAGTGGTGTCAAAATAACTGGTCACTTCTATGCTTG 300
|||||
QY 1888 ccaggttgccagagaaaaatccaggaaattctgactcaggtaaaagcagcaccaacaaca 1947
|||||
Db 301 CCAGTTGCCAGAGAAAAATTCNGGAAATTCGACTCAGGTAAGAGCAGCACCAACAACA 360
|||||
QY 1948 gaaggctctgcaaaagtggaccacctcagtcagaaagcagaaagtcaggagaaacagcc 2007
|||||
Db 361 GAAGGCTCTGCAAAAGTGACCACTCAGTCAAGACGGAAGTAAGAGGCTCAGGAAACAGCC 420
|||||
QY 2008 caccacagagcgagatgccaaacaaacagacagattgcttaaccaaacagatggcgctgac 2067
|||||
Db 421 CACCACAGAGCGAGATGCCAAACCAACAGACAGATTGCTTAACCAACAGATGGCGCTGAC 480
|||||
QY 2068 cccctaccagaatcacatgcacaagtttttaccctagcaggtgttctcagagaccagcc 2127
|||||
Db 481 CCCGTATCCAGAAATCAGATGCACAAGTTTTTACCCTAGCCAGTTGTTTCTGAGGACCCAGGC 540
|||||
QY 2128 aacttttgaaactcctctctgtgagaaatgtatacttttatgctctctctgaaatgtatgaca 2187
|||||
Db 541 AACTTTTGAACCTCTGTCTGTGAGAAATGTATACTTTATGCTCTCTGAAATGTATGACA 600
|||||
QY 2188 ccagcttttaaaacaaacaaacaaacaaacaaaggggtggggagggagggaaagag 2247
|||||
Db 601 CCAGCTTTTAAACAAACAAACAAACAAACAAAGGGTGGGGAGGGAGGAAAGAG 660
|||||
QY 2248 aagagctctgacactccctctgtgttagtctcacagataacagatatctctctct 2307
|||||
Db 661 AAGAGCTCTGCACTTCCTCTGGTGTAGTCTCAGGTATACAGATATATCTTAATCTTCT 720
|||||
QY 2308 taatatcccccaataatgcccagaaattggcttaagtgtcttccactaaattcatcaaat 2367
|||||
Db 721 TAATATTTCCCCCAATAATGCCAGAAATT-GCTTAATGATGCTTTC-CFAAATTCATCAAT 778
|||||
QY 2368 agattgctctaaatccaaattgttaaatggatcgagaataattatcacaggaaacttaa 2427
|||||
Db 779 AGATTGCTCTAAATCCAAATGGTAAATTTGGATCAGAAAT-ATTATCACAGGAACCTTAAA 837
|||||
QY 2428 tggtaagccattagcatagaaaaactgttc-tcagtttttttttttaccctaacactaacat 2486
|||||
Db 838 TGGTNAGCCCTTTGCATAGAAAACCTGGTCTTAAGTTTAAATTTTACCTAACCCCTTACAT 897
|||||
QY 2487 gagtaac 2493
|||||
Db 898 GAGTAAC 904
|||||
RESULT 3
AUI42785
LOCUS AUI42785 y79AA1 Homo sapiens cDNA clone y79AA1000871 5', mRNA 25-OCT-2000
DEFINITION sequence.
ACCESSION AUI42785
VERSION AUI42785.1 GI:11004306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS Ota,Y., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952

Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
source 1..876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="y79AA1000871"
/clone_lib="y79AA1"
/cell_type="retinoblastoma"
/cell_line="Y79"
/note="Vector: pME18SFL3"
BASE COUNT 289 a 181 c 183 g 220 t 3 others
ORIGIN
Query Match 19.2%; Score 802; DB 10; Length 876;
Best Local Similarity 97.8%; Pred. No. 4.4e-118;
Matches 854; Conservative 0; Mismatches 13; Indels 6; Gaps 4;
QY 1676 gaagaaaaactttgttagctcctaagaagaggtgaaccttaagctcatatcagagtacca 1735
|||||
Db 1 GAAGAAAACCTTTGTGTAGTCTCTAAAGAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCA 60
|||||
QY 1736 tcctttgctgctgcagagattattgaaaaggaggcaaaacggtgaatgaacttcagaat 1795
|||||
Db 61 TCCTTTGCTGCTGCGACAGATTATTGGAAAAGGAGGCAAAACGGTGAATGAACCTTCAGAAT 120
|||||
QY 1796 ttgtcaagtcagaaagttgtgttcctctgtgaccagacacctgtatgagaatgacaaagt 1855
|||||
Db 121 TTGTCAAAGTCAGAAAGTTGTTGCCCTCGTGACCAGACACCTGTATGAGAAATGACCAAGTG 180
|||||
QY 1856 gtgtcaaaaaaactggtcacttctatgtccaggtgcccagagaaaaaattcagaaa 1915
|||||
Db 181 GTGTCAAAAATAACTGTGCTACTTCTATGCTGTGCCAGGTGCCCCAGAGAAAAAATTCAGAAA 240
|||||
QY 1916 attctgactcaggttaagcagcagcacaacagaggtcctgcaaaagtggaccacctcag 1975
|||||
Db 241 ATTCTGACTCAGGTAAAGCAGCAGCAACAACAGAAAGCTCTGCAAAAGTGGACCACTCAG 300
|||||
QY 1976 tcaagacggaagtaaaaggtcaggaacacagcccaacacagagcgagatgccaacccaag 2035
|||||
Db 301 TCAAGACGGAAGTAAGGGCTCAGGAAACAGCCACACAGAGCGAGATGCCAAACCAAG 360
|||||
QY 2036 acagattgcttaaccaacagatggcgctgacccctctatccagaatcacatgcacaagtt 2095
|||||
Db 361 ACAGATTGCTTAACCAACAGATGGCGCTGACCCCTTATCCAGAAATCAGATGCACAAAGTT 420
|||||
QY 2096 tttaacctagccagttgtttctgaggaccaggcaaaacttttgaactcctgtctctgtgaaa 2155
|||||
Db 421 TTTACCTAGCCAGTTGTTTCTGAGGACCAAGCAACTTTTGAACCTCTCTCTCTGTGAGAA 480
|||||
QY 2156 tgtatacttatgctctctgaaaatgtatgacacccagcttttaaaaaacaaacaaacaa 2215
|||||
Db 481 TGTATACTTTATGCTCTCTGAAATGTATGACACCCAGCTTTAAACCAACAAACAAACA 540
|||||
QY 2216 acaaaaaaagggtggggaggagggaagagagctctgcaactcctctctgtttgttag 2275
|||||
Db 541 AAAAAAAGGGTGGGGAGGGAGGAGGAGAGAGAGAGAGACTCTGCACTTCCCTTTCTTTGTAG 600
|||||
QY 2276 tctcacagtaaacagatatcttaattcttcttaattcccccataatgccagaaattg 2335
|||||
Db 601 TCTCACAGTATAACAGATATTTCTAATTTCTTTAATATTTCCCCCATTAATGCCAGAAATTG 660
|||||
QY 2336 gcttaatgatgctttcaactaaattcatcaaatagatgtgctccttaaaatccaaattgttaaaa 2395
|||||
Db 661 GCTTAATGATGCTTTCATCTAAATTCATCAATAGATTGCTCTCTTAATTCATTCGTAATA 720
|||||
QY 2396 ttggatcagataaattatcacaggaacttaattgttttaagccattag-catagaaaaactg 2454
|||||
Db 721 TTGGATCAGAAATAATATATATACAGGAACCTTAATATGTTAAGCCATTAGCCTTAGAAAACTG 780
|||||

Qy	547	tttactagtcacgata	ggaggtggagagctgtgagcaagtgaacactgactcggaaac	606
Db	575	CCTCCTGGCTCAGTACGTACGTGAGAACTGTGAGCAAGTGACACTGAAAGTGAAG	634	
Qy	607	tgagttgtaaatgaacctattccagtaagaccaaagctagacaagcaactagacaact	666	
Db	635	AGCTGTGGTCAACGTCACTACTTCTTAACCGGAGAGCAGACAGGCAAGCTATCATGAAGCT	694	
Qy	667	gaatggatttcagttagagaatttcaccttgaagttagctatataccctgatgaatggc	726	
Db	695	AAATGGCCATCAACTGSGAGAACCATGCCCTGAAGGCTCTCTACATACCTGATGAGCAGAT	754	
Qy	727	cgccagcaaaaccccttcagacagcccagaggtcgccgggggcttgggcaagaggggctc	786	
Db	755	AACGCAAGGTCCTGAGAATGGGCGTCGTGGAGGGCTTTGGGTCTCGGGGCCAGGCCGGCA	814	
Qy	787	ctcagggcagggggtctccaggatccgatacccaagcagaaacccatgtgatttgctctgcgc	846	
Db	815	AGGGTCGCCGTGGCAGCAGGGGCTCCAGCCCAAGCAGCAGCCAGTGAGCATCCCTCTCCG	874	
Qy	847	cctgctggttccccacccaattgtgtgagccatcatagaaaaaaggctgcaccattcg	906	
Db	875	GCTCCTGGTCTACGCAGTATGTAGGCGCTATCATTTGCAAGGAGGGTGCACCATCCG	934	
Qy	907	gaacatcaccaaacagaccagctctaaatcgatgtccacgtaaaagaaatgcgggggc	966	
Db	935	AAACATCACAAAACAGACGAGTCCAAAATAGACGTGATAGGAAGGAGAAATGCGGGCGC	994	
Qy	967	tgctgagaagtcgattactatctctactcctgaaggcacctctcgcgcttgaagtc	1026	
Db	995	TGCGGAGAGGCCATCAGCGTGATTCACCCCTCGAAGGCTGCTCTCTCCGGTGCAGAT	1054	
Qy	1027	tattctggagattatgataaaggaagctcaagataataaattcacagaagatccccctt	1086	
Db	1055	GATCTTGAGATTATGCACAAGGAGCAAGGACACCAAAACGGCAGATGAAGTTCCCT	1114	
Qy	1087	gaagattttagctcataataaactttgttgagcgtcttattggtaaaaggaagaatact	1146	
Db	1115	GAAGATCCTGGCTCATAAACATTCCTGGGGGACTCATTTGGCAAGGAAGGGCGGAACCT	1174	
Qy	1147	taaaaaattgagcaagacacagacactaaaatcacgatatctccattgcaggaattgac	1206	
Db	1175	GAACAAGGTGGACGAGACACAGACGAAGATCACCATCTCATCGCTCCAGGACCTCAC	1234	
Qy	1207	gctgataataccgaacgcactattacagttaaaggcaatgttgagacatgtgcctaaagc	1266	
Db	1235	GCTCTATAACCCCTGAGAGGACCATCACTGTGAAGGGCGCCATTGAGAACTGTTGACGGC	1294	
Qy	1267	tgaggagagatcatagaataatcaggagctctatgaaatgatattgcttctatgaa	1326	
Db	1295	CGACAGAGGATCATGAAGAAGTTGCGAGAGGCTTTACGAGAACGACGTTGGCGGCATGAG	1354	
Qy	1327	tcttcaagcacatttaattcctggattaaatctgaeccttgggtctgttccccaccac	1386	
Db	1355	CTTGACGTCCACCTCATCCCTGGGCTTAACTGGCTGCTGTGAGTCTCTTCCCAGCTTC	1414	
Qy	1387	ttcagggatgccacctcccacgtcaggggcccccttcagccatgactcctccctaccgcga	1446	
Db	1415	ATCCAGGGCTGTCCCTCTCTCCAGCAGTGTCAACGGGGGCTGCTCCTATATAGCTCCTT	1474	
Qy	1447	gtttgagcaatcagaacaggagactgttcacagttaattccagctctatacagtcggtgc	1506	
Db	1475	CATCGAGGCTCCGGAGGAGGATGTACAAGTGTTCATCCCGCCCGCAGGCTGTGGCGC	1534	
Qy	1507	catcatcggaacgagggcgacacatcaagcagcttctctcgctttgtctggaggttcaat	1566	
Db	1535	CATCATTTGGCAAGAGGGCCAGCACATCAACAACACTCTCCCGCTTGCCAGCGCTCCAT	1594	
Qy	1567	taagattgtccagcggaagcaaccagatgtctaaatgaggtgaggtattatcaactggacc	1626	
Db	1595	CAGAGTTGACCAACCAAGAAACACCTGATTCAAAAGTTCGAATGCTGCTCACTAGTGAAC	1654	

QY	1627	accagaggtcagttcaaggctcagggaagaatttaaggaaaaattaaagaagaaactt	1686
DB	1655	CCGAGAGGTCAGTTCAAGCGCCAGGGAAGAATCTATGGCAAACTAAAGAAGAGAAATTT	1714
QY	1687	tgtagtcctaaagaagaggtgaaacttgaagctcatatcacagagtgccatccttggctgc	1746
DB	1715	CTTTGGTCCAGAGGAGGAAGTAAAGCTAGAGACCCACATACAGGGTTCCGGCTTCAGCAGC	1774
QY	1747	tgccagaggtatttgtaaaagaggagccaaacggtgaatgaacttcagaatttgcaggtgc	1806
DB	1775	CGGCGGTGTCATCGGCAAGCGGCAAAACGGTGAATGAGCTGCAGAACTTGACCCGAGC	1834
QY	1807	agaagttgtgtccctcgtgaccagacacctgatgagaatgacccaagtgtgtcctcaaat	1866
DB	1835	TGAGGTGGTGTAGTCCCAAGAGAGACAGACCCCGGATGAGAACGACCAAGTCATTCTTTAAGAT	1894
QY	1867	aactggtcacttctatgcttccaggttcccagagaaaaattcagaaaattctgactca	1926
DB	1895	CATCGGACATTTCTATGTCAGCAGCATGCTCAGCGGAAGATCCGAGACATCTCGTCTCA	1954
QY	1927	ggtaagcagcagccaaacaag	1948
DB	1955	ACTTAAGCAACAGCACCAGAAG	1976
RESULT	5		
AU117427			
LOCUS	AU117427	808 bp	mrna
DEFINITION	AU117427	HEMBA1 Homo sapiens	EST
ACCESSION	AU117427	HEMBA1001372	19-OCT-2000
VERSION	AU117427	1	
KEYWORDS	AU117427	GI:10932395	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 808)		
	Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,		
	Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and		
	Isogai, T.		
TITLE	HRI human cDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isogai		
	Genomics Laboratory		
	Helix Research Institute		
	1532-3 Yana, Kisarazu, Chiba 292-0812, Japan		
	Tel: 81-438-52-3951		
	Fax: 81-438-52-3952		
	Email: genomics@hri.co.jp		
	HRI human cDNA project: 5' - & 3' - end one pass sequencing: Helix		
	Research Institute; cDNA library construction: Department of		
	Virology, Institute of Medical Science, University of Tokyo, and		
	Helix Research Institute.		
FEATURES	Location/Qualifiers		
source	1. .808		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HEMBA1001372"		
	/clone_lib="HEMBA1"		
	/tissue_type="whole embryo, mainly head"		
	/dev_stage="embryo, 10 weeks"		
	/note="Vector: pME18SFL3"		
BASE COUNT	262 a 127 c 160 g 255 t		
ORIGIN	4 others		
Query Match	18.2%	Score 759.2;	DB 10; Length 808;
Best Local Similarity	97.1%	Pred. No. 2.9e-111;	
Matches 782;	Conservative	0; Mismatches 21;	Indels 2; Gaps
QY	2367	tagattgctcctaaatcccaattgttaaaattggatcagaataattatcacaggaacttaa	2426

Db	1	TAGATTGCTCTTAATCAATTTGTTAAATTTGGATCAGAATAATATACAGGAACCTTAA	60
Qy	2427	atgttaagccattagcatagaaaaactgtctcagtttttatttttacctaacaactaaacat	2486
Db	61	ATGTTAAGCCATTAGCATAGAAAAACTGTTCTCAGTTTATTTTACCTAACACTTAACAT	120
Qy	2487	gagtaaacctaaggaagtcgaatggtgtgttgccaggggtattaaacgtgcatittttact	2546
Db	121	GAGTAACCTAAGGGAAGTGCATAGTGTGTGGCAGGGGTATTAAACGTGCAATTTTACT	180
Qy	2547	caactcacctcaggtatttcagtaatacaaatgaaaagcaaaattgtctctttttttgaaaa	2606
Db	181	CAACTACCTCAGGTATTTCAGTAATACATGAAGAAGCAAAATGTTCCTTTTTTTGAAAA	240
Qy	2607	ttttatacttttataatgatagaagtcacacccgtttttttaaataaaattttaaatttt	2666
Db	241	TTTTATATACTTTATAATGATAGAAGTCCAACCCGTTTTTTAAAAAATAAATTTTAAATTT	300
Qy	2667	aacagcaatcagctaaacaggcaaatgaagatttttactcttgctggtgagcagtaaaagct	2726
Db	301	AACAGCAATCAGCTAACAGGCAAAATTAAGATTTTACTTCTGGCTGTGACAGTAAAGCT	360
Qy	2727	ggaaaattcaatttcagggttttttgaggtcttttgacacagttatttagttaaatcaaatgt	2786
Db	361	GGAAATTAATTTTTCAGGGTTTTTTGAGGCTTTTGACACAGTATTAGTTAAATCAAAATGT	420
Qy	2787	tcaaaaaacagcgagcagtcgcttagtatctggagagcagcactaccattttctttcatt	2846
Db	421	TCAAAAATACGAGCAGTGCCTAGTATCTGAGAGCAGCAGTACCATTATTTCTTTCATT	480
Qy	2847	tatagttgggaaagtttttgacggtactaacaagaagtgctgcagagagattttgaaacggc	2906
Db	481	TATAGTTGGGAAAGTTTTTGTACGGTACTTAACAAAGTGTGTCAGGAGATTTTGGAAACGC	540
Qy	2907	tggtttaaatgcctcaagagacitcagttttttgtttgttagctacatgattgaatgcataa	2966
Db	541	TGGTTTAATGGCTTCAGGAGACTTCAGTTTTTGTGTTAGCTACATGATTGAATGCATAA	600
Qy	2967	taaatgttttgcctctgactatcaatcacctaaagaagtgcatcagtgagagatgca	3026
Db	601	TAAATGCTTTTGCTCTGACTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA	660
Qy	3027	agactttcaactgactggtgcaaaagcaagcttttagcttgttcttagtagatgcttagtttg	3086
Db	661	AGACTTTCAACTGACTGTGCAAAAGCAAGCTTTAGCTTTGGCTTATAGGATGCTTAGTTTG	720
Qy	3087	ccactacattcagacaatggagcagtcataagtggtgtgacagtggttttaaac--gcaa	3144
Db	721	CCACTACACTTTAGACCAATGGGACAGCATANATGTGGGGGGTGGGGTTTAAACGCCAC	780
Qy	3145	caaaaggctacatttccatggggcc	3169
Db	781	CAAAAGGCTACATTTCCATGGGGC	805
RESULT	6		
AUI34323			
LOCUS	AUI34323	745 bp	mRNA
DEFINITION	AUI34323	OVARC1 Homo sapiens	cdna clone OVARC1001723 5', mRNA
ACCESSION			sequence.
VERSION	AUI34323		
KEYWORDS	AUI34323.1	GI:10994862	
SOURCE			EST.
ORGANISM			human.
REFERENCE			Homo sapiens
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
			1 (bases 1 to 745)
			Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
TITLE			HRI human cdna project
JOURNAL			Unpublished (2000)

COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cdna project; 5', - & 3'-end one pass sequencing: Helix Research Institute; cdna library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES	Location/Qualifiers
source	1..745 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="OVARC1001723" /clone_lib="OVARC1" /tissue_type="ovary, tumor tissue" /note="Vector: pME18SFLJ"
BASE COUNT	232 a 168 c 170 g 172 t 3 others
ORIGIN	
Query Match	17.3%; Score 723.6; DB 10; Length 745;
Best Local Similarity	98.9%; Pred. No. 1.4e-105;
Matches 737; Conservative	0; Mismatches 7; Indels 1; Gaps 1;
Qy	1455 aatcagaacccagagactgttcacagtttattccagctctatccagctctatcagtcggtgccatcagc 1514
Db	1 AATCAGAAACGGAGACTGTTTCATCTGTTTATCCAGCTCTATCAGTCGTGTCATCATCG 60
Qy	1515 gcaagcagggccagcacatcaagcagctttctcgcttctgctgagcttcaattaaagtgg 1574
Db	61 GCAAGCAGGGCCAGCACATCAAGAGCTTTCTCGCTTTGCTGAGCTTCAATTAAGATTG 120
Qy	1575 ctccagcggaagcaccagatgctaaaagtgaggatgagtgattatcactgagcaccagaggg 1634
Db	121 CTCACGCGGAGCAGACAGATGCTAAAGTGAGGATGGTATTATCAGTCGTGACCACCCAGAGG 180
Qy	1635 ctcaagttcaagctcagggagaattttatgaaaaattaaagaagaaactttgttagtc 1694
Db	181 CTCAGTTCAAGGCTCAGGGAAGAAATTTATGGAAAAATTAAGAAAGAAACTTTGTTAGTTC 240
Qy	1695 ctaaaagagaggtgaaacttgaaagctcatatcagagtgccatccttttgcgtgagcagag 1754
Db	241 CTAAAGAGAGGTGAAACTTTGAAGCTCATATCAGAGTGCCATCTTTGCTGTGGCAGAG 300
Qy	1755 ttattggaaaaagggcaaaacgggtgaatgaaacttcagaatttgcgaagtgcagaagtgg 1814
Db	301 TTATTGGAAAAAGGAGGCAAAACGGTGAATGAATCTCAGAAATTTGTCAAGTCGAGAAGTTG 360
Qy	1815 ttgtccctcgtagccagacacctgatgagaatgaccaagtgtgtcaaaataaactggttc 1874
Db	361 TTGTCCTCGTGACCAGACACCTGATGAGAAATGACCAAGTGGTGTGTCAAAAATAACTGTGTC 420
Qy	1875 actctatgtctgcccaggttgcccagagaaaaattcagaaaaattctgactcaggtgaaagc 1934
Db	421 ACTTCTATGCTTGCCAGGTTGCCAGAGAGAAAAATTCAGGAAATTCGACTCAGGTAAAGC 480
Qy	1935 agcaccacaacagaagcgtctgcaaaagtggaccacctcagtcgaagacggaagtaaaagc 1994
Db	481 AGCACCAACACAGAAAGGCTCTGCAAAAGTGGACCACTCAGTCAAGACGGAAGTAAGGC 540
Qy	1995 tcagaaaaacagccaccacagagcgagatgccaaacacaaagacagattgtcttaaccaaca 2054
Db	541 TCAGGAACAGGCCACACACAGAGGAGATGCCAAACCAAGACAGATGCTTAAACCAACA 600
Qy	2055 gatggcgcgtgacccctcattccagaatcactgcacaagtcttttactgacgagctgtttt 2114
Db	601 GATGGCGCTGACCCCTATCCAGAAATCATCATGACAAAGTTCCTACCTAGCCAGTGTGTT 660
Qy	2115 ctgaggaccagggcaacttttgaaactcctctctctgtgtgagaatgtatacctttatgtctct 2174

Db 561 CTGAGGACACGCACTTTTGAACCTCCTGCTGTGTGANAATGTAT-CTTTATGCTCTCT 719

Qy 2175 gaaatgtatgacacccagcttttaa 2199

Db 720 GAAATGTATCACNCCCACITTAATA 744

RESULT 7

LOCUS BE869656

DEFINITION BE869656 902 bp mRNA EST 20-OCT-2000

601446879F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850963 5', mRNA sequence.

ACCESSION BE869656

VERSION BE869656.1 GI:10318536

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM9570 row: n column: 20
High quality sequence stop: 683.

FEATURES

source Location/Qualifiers

1..902

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3850963"

/clone_lib="NIH_MGC_65"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 297 a 187 c 210 g 208 t

ORIGIN

Query Match 17.1%; Score 716.2; DB 11; Length 902;

Best Local Similarity 92.7%; Pred. No. 1.8e-104;

Matches 776; Conservative 0; Mismatches 53; Indels 8; Gaps 2

Qy 1509 tcacgcgcagcaggccagcacatcaagcagcttctcgtttgtcggagcttcaatta 1568

Db 1 TCATCGCGAAGCAGGCGCCAGCATCAAGCAGCTTCTCGTCTTGTGGAGCTCAATTA 60

Qy 1569 agattctccagcgcagcagcaccagatgctaaagtgcgagtggtgattatcactggaccac 1628

Db 61 AGATTGCTCCAGCGAAGCAGCAGATGCTAAAGTGAGGATGGTGATTTATCACTGGACCAC 120

Qy 1629 cacagctcagttcaagctcaggaagaattatggaataattaaagaagaaactttg 1688

Db 121 CAGAGGCTCAGTTCAGGCGCTCAGGGAAGAATTATGGAATAATTTAAAGAAGAAACTTTG 180

Qy 1689 ttagtctctaaagaaggtggaacttgagctcatatcagagtgccatcctttgctgctg 1748

Db 181 TTAGTCTCTAAAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTTGTGCTG 240

Qy 1749 gcagagctattggaagagggcgaacacgg tgaatgaacttcagaattgtcagaatgcag 1808

Db 241 GCAGAGTTATTGGAAGAGGCGCAACCGTGAATGAACCTTCAGAACTTCAGAACTTGTCAAGTGCAG 300

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```
BASE COUNT      288 a   159 c   159 g   331 t
ORIGIN

Query Match      15.5%; Score 648.4; DB 11; Length 937;
Best Local Similarity 90.0%; Pred. No. 1.1e-93;
Matches 777; Conservative 0; Mismatches 71; Indels 15; Gaps 8;

QY 3237 aaaaaaattagactccacttaagttagtaagtagtaacaggaattctctatactgtgcaa 3296
Db 1 AAAAAAATTAGACTCCACCTTAAGTAGTAAAGTATAACAGGATTTCTGTATACCTGTGCAA 60

QY 3297 tcagtcttttaaaaaaagtcacaaagtagagaatacaagaaaaagtttttngggataaa 3356
Db 61 TCAGTTCCTTTGAAAAAAGTCAAAGATAGAGAAATACAAGAAAAAGTTTGGGGATATAA 120

QY 3357 ttgaatgactgtgaaacatatgacctttgtataacgaactcattgtctcactcttgac 3416
Db 121 TTTGAATGACTGTGAACATATGACCTTTGATAACGAACTCATTTGGCTCACTCCTTGAC 180

QY 3417 agcaagcccgactacgtacaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3476
Db 181 AGCAAGCCCACTAGTACAAATGTTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 240

QY 3477 tctgaattgatttttttgattttgttgnatgnaagatgatacagcnaatgtttacactgatct 3536
Db 241 TCTGAATTGATTTTGTGAGTTT- GTTTGTGAAGATGATCACAGTCATGTTACACTGATCT 299

QY 3537 tnaagggacatatnttatacccttttaaaaaaaacccctgcctcattcttatttcgag 3596
Db 300 RAAGG- --ACATATATATAACCCTTAAAAAATAATCAGTGCCTCATCTTATTTCAAG 356

QY 3597 atgaatttcgatacagactagatgtcttcttgaagatcaatagacatntgaaatgat 3656
Db 357 ATGAATTTCTATACAGCTAGATGTTTCTGAAGATCAATTAGACATTTTGAATAATGAT 416

QY 3657 ttaagagtcttcttaagtctctgaaacaaagttctttttagttttaaccacaaaaa 3716
Db 417 TTAAGTGTTCCTTAAGTTTCTCTGAAAAACAAGTTTCTTTGTAGTTTAAACCAAAA 476

QY 3717 agtgcctttttgtcactgtgttctcctagcattcatgatatttttttccacacaaatgat 3776
Db 477 AGTGCCTTTTGGTCACGTGGATCTCCTAGCATTCATGATTTTTTTTCATACAAATGAT 536

QY 3777 taaaattgctaaaaatcatggaactgttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3833
Db 537 TAAAAATGGCTAAAAATCATGGAGTGGCTTCTGTGTGGATTTCCAGGTAACGATGTGTT 596

QY 3834 aagccagagcttttctcagattt- gattttttcccccaatattgtattttttaaaaaat 3892
Db 597 AGCGCCAGAGCTTTTCTCAGTATTTTCGATTTTTTTTCCCAATATTTGATTTTTTAAAAAT 656

QY 3893 aTacaCaggaagctgtcatttaaaacccctgctgttttaattctgcanaatttcacttcta 3952
Db 657 ATACACATAGGTGCTGCATT- -ATATCTGCTGTTTAAATTCGTGCATATTTCACTTTCTA 714

QY 3953 gcccttttagtatggcnaatcanaatttacttttaacttaagcaatttggtaa--tttggagta 4010
Db 715 GCCTTTAAGTATGGCAAAATCATATTTACTTTTACTTAAAGCTATCTGACTCTGGGAGAA 774

QY 4011 tctgtactagctaagaaa--taattcnataattgagttttgttactcncnaaanaatgggt 4068
Db 775 TCTGTACTAGCTAAGAAAAATAACTCCCTATTATTAGTGTGGTGACTCACTCACTATATGAT 834

QY 4069 cattctcatgnataatgttnc 4091
Db 835 C-TTCTCTCTGTATATGTGCC 856
```

```
RESULT 9
LOCUS   AUI27673
DEFINITION AUI27673 NT2RP2 Homo sapiens cDNA clone NT2RP2001818 5', mRNA
sequence.
ACCESSION AUI27673
VERSION   AUI27673.1 GI:10988027
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
           Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
           Isogai,T.
TITLE      HRI human cDNA project
JOURNAL   Unpublished (2000)
COMMENT    Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3951
           Fax: 81-438-52-3952
           Email: genomics@hri.co.jp
           HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
           Research Institute; cDNA library construction: Department of
           Virology, Institute of Medical Science, University of Tokyo, and
           Helix Research Institute.
FEATURES  Location/Qualifiers
           source
             1..733
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="NT2RP2001818"
               /clone_lib="NT2RP2"
               /cell_type="teratocarcinoma"
               /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
               cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 212 a 181 c 148 g 180 t 12 others
ORIGIN
```

```
Query Match      15.3%; Score 638.8; DB 10; Length 733;
Best Local Similarity 96.5%; Pred. No. 4e-92;
Matches 670; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

QY 826 accatgtgattgctctgcctgcctgcctgtgtcccccacaaatttgttggagccatcagg 885
Db 1 AACATGTGATTTTGCCTCTGCGCTGCTGTGTTGCCCAATTTGTTGGAGCCATCATAGG 60

QY 886 aaaaagaagtgcaccattcgaaacatcaccaaacagaccagctcaaaatcgatgtcca 945
Db 61 AAAAGAAGTGGCCATTTCGGAACATCACCAAAACAGACCCAGCTCAAAATCGATGTCCA 120

QY 946 ccgtaaaagaaatgcggggctgcgtgagtcgattactctctactctactctgaagg 1005
Db 121 CCGTAAAGAAATGCGGGGGCTGCTGAGAAGTCGATTACTATCTCTCTACTCTCTGAAGG 180

QY 1006 caccctgcgggttgtaagtcattctattcttgagattatgcataaggagctcaagataaaa 1065
Db 181 CACCTCTGCGGCTGTGAAGTCTATTCTGGAGATTATGTCATAGGAAGCTCAAGATATAA 240

QY 1066 attcagaagagatccctctgaagattttagctcataataactttgttggagctcttat 1125
Db 241 ATTACANAAGATACCCCTTGAAGATTTTACCTATAATAACTTTGTGGAGCTTTAT 300

QY 1126 tggtaagaagaaagaaaattttaaaaaaattgagcaagacagacagactaaaatcacgat 1185
Db 301 TGGTAAAGAAGGAGAAATCTTAAAAAANTTGAGCAAGACACAGACNCCTAAAAATCAGAT 360

QY 1186 atctccattgcaggaattgacgctgtataatccagaacgcactattacaggttaaaagcca 1245
|||||
```

```

Db 361 ATCTCCATTGCGAATTGACGCTGTATATATCCAGACGCGACCTATTACAGCTTAAGCGCAA 420
QY 1246 tgttgagacatgtgccaaagtgcaggaggagatcatgaagaataacaggagtgctttatga 1305
Db 421 TGTGTGACACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATACAGGCGAGCTTTATGA 480
QY 1306 aaatgatattgcttctatgaattcattcaagcacacatttaattcctcgttgattaaatctgaacgc 1365
Db 481 AAATGATATTGCTTCTATGAATCTTCNAGCACATNTAATTCCTGNATPTAAATCTGAACGC 540
QY 1366 cttaggtgctgttcccccacttcaggatgccacctccacactcaggccccccttcagc 1425
Db 541 CTGCGGTCTGTTCCACCCACTTCAGGGATGCCACCTCCACCTCAGGCGCCCCCTTCAGC 600
QY 1426 catgactcctccctaccgcagtttgagcaatcagaacaggagactgttcatcagtttat 1485
Db 601 CATGACTCCTCCCTNCCGCGAGTTTGAGCAATCCGAAACGAGACGTGTTTCNTCTGTTT 660
QY 1486 ---cccagctctat-cagtcgggtgccaatcagg 1515
Db 661 ATCCCCACCTCTATCCAGTCGGTGGCCATCATCCG 694

RESULT 10
AUI39887 751 bp mRNA EST 25-OCT-2000
LOCUS AUI39887 PLACE1 Homo sapiens cDNA clone PLACE1011527 5', mRNA
DEFINITION sequence.
ACCESSION AUI39887
VERSION AUI39887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 751)
Ota.T., Nishikawa.T., Suzuki.Y., Ishii.S., Saito.K., Kawai.Y.,
Yamamoto.J., Wakamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and
Isogai.T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
Source 1..751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1011527"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SFL3"
BASE COUNT 253 a 131 c 137 g 226 t 4 others
ORIGIN

Query Match 15.2%; Score 634; DB 10; Length 751;
Best Local Similarity 94.5%; Pred. No. 2.3e-91;
Matches 718; Conservative 0; Mismatches 30; Indels 12; Gaps 6;

QY 2977 gtgcttctgactatcaataacctaagaagtgcatcagtgaaagatgcaagactttcaa 3036
Db 1 GTGCTCTTGACTATCAATACCTAAAAAAGTGCAATCAGTCAAGAGATGCAAGACTTTCAA 60
QY 3037 ctgactggcacaagaacgcttagctgtcttcttataggatgcttagttgccaactacact 3096
```

```

Db 61 CTGACTGGCAAAAGCAAGCTTTAGCTTGTCTTATAGATGCTTTAGTTGCCATACACT 120
QY 3097 tcagaccaatgggacagctcatagatgggtgacagtggttttaaacgcacaaagagcttaca 3156
Db 121 TCAGACCAATGGGACAGCTATAGATGGTGTGACAGTGTTTAAAGCGCAACAAAGGCTACA 180
QY 3157 ttccatggggccagcactgcatgagccctcaactaagctattttgaagatttttaagcac 3216
Db 181 TTTCCATGGGCGCAGCACTGTCATGAGCCTCCTCAAGCTATTTTGAAGATTTTAAAGCAC 240
QY 3217 tgataaataaaaaaataaaataaataaataaataaataaataaataaataaataaataaag 3276
Db 241 TGATAAAT-----AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAG 294
QY 3277 gattctgtatactgtgcaatcagttcttgaaaaaaagctcaaaagatagagaatacaa 3336
Db 295 GATTTCTGTATATCTGTGCAATCAGTCTCTTTGAAAAAATAAAGTCAAAAGATAGAGAA 354
QY 3337 gaaaagtttngggatataaatttgaatgactgtgaaacatatgacctttgataaacgac 3396
Db 355 GAAAAGTTTTTGGGATATAAATTTGAATGACTGTGAAAAACATATGACCTTTGATAACGAC 414
QY 3397 tcatttgcactcctctgacagcaaaagccagtcagtcacaaattgtgttggttggtgg 3456
Db 415 TCATTTGCTCCTCCTCTTGACAGCAAAAGCCAGTACAGTACAAATTTGTTGGGTGGGTGG 474
QY 3457 tctccaagggccagcgtctctgaaattgttttggagtttggnttgnnaagatgatca 3516
Db 475 TCCTCAAGGCGACGCTCTCTCTGAATGATTTTGGAGTTTTT-GTTTGTGAAGATGATCA 533
QY 3517 cagncatgttacactgatctttnaagacatatntataaccctttaaataaaataacccc 3576
Db 534 CAGTCATGTTTACACTGATC-TAAAGGACATATA-TATAACCTTTTAAAAAAGAAATCAC 591
QY 3577 tgctcattcttatttgcagatgaatttcgatataca-gactagatgtcttctgaagatca 3635
Db 592 TGCTCATCTCTTTTCAAGATGAATTTCTATACGAGCTAGATGTTTTCGGAAGAACA 651
QY 3636 attagacattntgaaatgatttaaagtggttttcccttaattctctgtaaaacaagtttc 3695
Db 652 ATTAGACATTTTCAAAATGATTTAAAGTGTTTTTCCCTTAANGTCTCCGAAACAAGTTC 711
QY 3696 --ttttagtatttaaccccaaaagtcacctttttgtcac 3733
Db 712 CNTTTGGAGGTTTAAACCCAAAAAAGTGCCCTTTTGNCAAC 751

mRNA sequence.
BE545535 790 bp mRNA EST 09-AUG-2000
601070391F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456407 5',
ACCESSION BE545535
VERSION BE545535.1 GI:9774180
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 790)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

Plate: LLM8444 row: f column: 24
High quality sequence stop: 667.

FEATURES

```

Location/Qualifiers
1..790
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3456407"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="Drl10B"
/note="organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

```

BASE COUNT	221 a	193 c	209 q	167 t
communities.				

ORIGIN

Query Match	15.1%;	Score 630.4;	DB 10;	Length 790;
Best Local Similarity	93.2%;	Pred. No. 8.3e-91;		
Matches 714;	Conservative	0;	Mismatches 46;	Indels 6;
				Gaps 5;

Qy 333 cgggtgcgggacccttcctgggtgaagactggctacgcgttcgtggaactgcccggacqaga 392

[illegible]

Qy 393 gctggccctcaaggccatcgaggcgcttccaggtaaaatagaactgcacgggaaaccca 452

61

501 GCATGGGCCATCGAGGCGCTTTCAGGTAAAAATAGAACTGCACGGGAACCCA 120

Qy 453 tagaagttgagcactcgggtcccaaaaggcaaggattcggaaacttcagatacgaata 512

Db 121 TAGAAGTTGAGCAGTCGGTCCCAAAAGCCAAAGCATTCGCCAAACTTACCAATA 180

[illegible]

Qy 513 tccgcctcatttacagtgagggtgctggatagtttactagtagtccagtagggagtgg 572

Db 181 TCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTACTAGTCCAGTATGCAGTGGTGG 240

.....

Qy 573 agagctgtgagcaagtgaacactgactcggaactgcagttgtaaatgtaacctattcca 632

Db 241 AGAGCTGTGAGCAAGTGAACACTGACTCGGAACTGCAGTTGTAATGTAACTATTCCA 300

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

053 gcaaggaccbaagcclayacaaagccacclagacaaactgaatgatttcagttagagaatttca 692

Db 301 GTAAGGACCAAGCTAGACAAGCACTAGACAACTGAATGGATTTCAGTTAGAGAATTCA 360

0v 693 ccttgaagtaccttatatccctgatatgaataagcccccaaacacccct+accacccc
753

[illegible]

Db 361 CCTGAAAGTAGCCTATATCCCTGATGAAATGGACGCCAGCAAAACCCCTTGCAGCAGC 420

Qy 753 cccgaqtcqccqggqgcttggqccagaggggctcctcaaggcagggggtctccaggatcca 812

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 421 CCCAGGTCGCCGGGGGCTTGGGCAGAGGGGCTCCTCAAGGCA-GGGTCTCCAGGATCCG 479

Qy 813 tatcaagcagaaaccatgtgattgcctctgcgcctgctggttccaccaatttqttg 872

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

DB 480 TATCCAAGCAGAAACCATGTGATTTGGCTCTGTGCGCGTGTGGTTCCACCAATTTGTGG 539

QY 873 gagcatcata-ggaaaagaagggtgccaccattcggaacatc-accaaacagaccagtc 930

```
Db 470 AAAAAATTAATTTAAATTTAACAGCAATCAGCTAACAGGCAAAATTAAGATTTTTACTTCT 529
QY 2708 ggctggtagacagtaagctggaaaaataatttcagggttttttgaggcttttgacacagt 2767
Db 530 GGCTGGTGACCGTAAGCTGGAATAATTAATTCAGGGTTTTTTGAGGCTTTTTCACACAGT 589
QY 2768 tattagttaaatacaaatgtttcaaaaatacaggagcagtcctagtagtatctgagagcgacac 2827
Db 590 TATTAGTTAAATCAAAATGTTTCAAAAATACGAGCAGTGCCTAATATCTGGAGAGCAGCAC 649
QY 2828 taccattattcttccatttat-agttgggaagattttgacgggtactaacaagaagtggt- 2885
Db 650 TACCAATTATCTTCTTATTAAGTTGGGAAAGTTTTTGCACCGTACTAACAAAGTGGTC 709
QY 2886 cgcagagattttggaacggctggtttaaatggcttc-agagagacttcagttttttgttt 2944
Db 710 CGCAGGAGATTTTGGACGGCTGGGTAAATGCTTCAAGGAGACATTCAGTTTTTTGGTT 769
QY 2945 agc--tacatgattgaatgcataataaaatgctttgtg 2979
Db 770 ANNCTACAATGATTGAATGCATATAATAATGCNTTGGG 806

RESULT 13
BG108548 805 bp mRNA EST 30-JAN-2001
LOCUS 602278519f1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4366085 5',
DEFINITION mRNA sequence.
ACCESSION BG108548
VERSION NIH_MGC.1 GI:12602394
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://limage.llnl.gov
Plate: LLAM10016 row: f column: 06
High quality sequence stop: 722.
FEATURES
Source Location/Qualifiers
1..805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4366085"
/clone_lib="NIH_MGC_86"
/tissue_type="Osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bone; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 242 a 130 c 136 g 297 t
ORIGIN

Query Match 14.8%; Score 617.6; DB 11; Length 805;
Best Local Similarity 90.9%; Pred. No. 8.9e-89;
Matches 728; Conservative 0; Mismatches 62; Indels 11; Gaps 7;

QY 3280 ttctgtactgtgcacatcagttcttggaaaaaagtcacaaagatagagaata-caaga 3338
|||||
```

```
Db 5 TTCTGTATACTGTGCAATCAGTTCTTTTGAAAAAAGCTAAAAAGATAGAGAAATACCAGA 64
QY 3339 aaagttttinggataataatttgaatgactgtgaaacatatgacccctttgataacgaactc 3398
Db 65 AAAGTTTTTTGGGATATAAATTTGAATGACTGTGAAAAACATATGACCTTTTGATAAGCACTC 124
QY 3399 atttgcctaccccttgacagcaaaagcccgactacgtacaattggttgggtggtgggtgc 3458
Db 125 ATTTGCTCACTCTTGCACAGCAAGCCCGACGTACGTACAATTTGTTGGTGGTGGTGGTGC 184
QY 3459 tccaaagccacgcgtcctctgaattgatttttttgaatttttgaattttgnttgnagaagatcacaa 3518
Db 185 TCCAAGGCCACGCTGCTCTCTGAATTGATTTTTTTGAGTTT - GTTTGTAGATGATCACA 242
QY 3519 gncatgttacactgatctttnaaggacatatntatatacccttttaaaaaaaatccctgt 3578
Db 243 GTCATGTACACTGATCTAAAGG --ACATATATATAACCCCTTTAAAAAATACTACTG 299
QY 3579 cctcattcttatttcagagatgaatttcogatacagactagatgcttctctgaagatcaatt 3638
Db 300 CCTCATTTCTATTTCGAAGATGAATTTCTATACAGACTAGATGTTTCTCGAAGATCAATT 359
QY 3639 agacattntgaaaaatgatttaaaagtgttttcccttaattgtctctctgaaacaaagtctctt 3698
Db 360 AGACATTTTGAATAATGATTTAAAGTGTTTTCCCTTAATGTTCTGTGAAAAACAAGTTTC-TT 418
QY 3699 tglagttttaaccacaaaagtgccctttttgtcactgggtttctcctcctagcattcatgattt 3758
Db 419 TGTAGTTTTAACCAAAAAGTGGCC -TTTGTGTCACCTGGATTCCTCCPAGCATTTCATGATT 477
QY 3759 tttttccacacaatgaattaaaaattgctaaaaatcagtgactggcttctctgggtggatttc 3818
Db 478 TTTTTCATACAATGAATTAATAATTCGTAATAATCATGGACTGGCTTCTGTTGGATTTC 537
QY 3819 agctaagatgtgtttaagccagagcttttctcagatttgcattgattttttccccaatttt 3878
Db 538 AGTAAGATGTGTTTTAAGGCCAGAGCTTTTCTCAGATTATTGATTTTTTCCCCCAATATTT 597
QY 3879 gatttttataacacacacacacacacacacacacacacacacacacacacacacacacacac 3938
Db 598 GATTTTAAAAAATATACACATAGGTGCTGCATTT-ATATCTGCTGCTGCTTTAAATTCCTGTC 656
QY 3939 anatttcactcttagccttttagtagtggnnaatcaanaatttacttttacttaagcatttg 3998
Db 657 ATATTTCATCTTAGCTTTTGTAGTGGCAATCATATCTTACTTTTACTTAAGCATTT 716
QY 3999 taattggagatctgctactagcagaataaattcnaataattgatttgcactcnc 4058
Db 717 GTAATTGGAGTATCTGCTAGCTAGGCTAAGAAATTAATCATATAAATGGAG--TTGTACTCACC 774
QY 4059 aanatgggtcattcctcatg 4079
Db 775 ATATATGGATCATTTCCCGAGG 795

RESULT 14
BF608004 769 bp mRNA EST 01-APR-2001
LOCUS MYL_000946 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
DEFINITION CDNA clone ICRFp522M2029 5', mRNA sequence.
ACCESSION BF608004
VERSION BF608004.1 GI:13504496
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 769)
Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,
Lehrach,H. and O'Brien,J.
Detection of a high number of novel genes in a 9-day mouse embryo
CDNA library normalised by oligonucleotide fingerprinting
Unpublished (2001)
JOURNAL
```

COMMENT

Contact: Hennig S
Laboraty 123, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
EST's are made from clones being representatives of clone clusters.
Clone clusters were calculated from oligonucleotide fingerprints.
PCR Primers
FORWARD: 5'-GAGCTATTCAGAGTACTGA-3'
BACKWARD: 5'-TAATAGCAGTACTATAGG-3'
Seq primer: 5'-ATTAGGTGACACTATAG-3'
High quality sequence stop: 769.
Location/Qualifiers
1. .769

FEATURES

source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="ICRFp522M2029"
/clone_lib="Mouse 9-day fetus cDNA library ICRFp522"
/tissue_type="whole embryo"
/dev_stage="embryonic 9-day"
/lab_host="E.coli, XL1 blue"
/note="Vector: PVSport1; Site_1: NotI; Site_2: SalI;
Library preparation by oligo-dT priming of RNA. Clones can
be ordered from the Resource Center in Berlin,
http://www.rzpd.de."
208 a 195 c 191 g 174 t 1 others

Query Match 14.8%; Score 617; DB 11; Length 769;
Best Local Similarity 91.5%; Pred. No. 1.le-88;
Matches 664; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 1195 gcaggaattgacgtgtataatccagaacgcactattacagttaaggaatgttgagac 1254
|||||
Db 44 GCAGGAACGTGCGTGTACATCCGGAACGCCACCATTCAGTGAAGGCGTGTGAGAC 103
Qy 1255 atgtccaaaactgaggagatcataagaaaaatcagggagctttatgaaatgat 1314
|||||
Db 104 GTGTCCCAAGCGGAGGAGGAATAATGAAGAAGATCAGGAGCTTTATGAAATGATAT 163
Qy 1315 tgcctctatgaattccaagcacattatctctgattataatctgaacgccttgggtct 1374
|||||
Db 164 TGCTTCCATGATCTTCAGGCACATTTAATCCCTGGATTAATCTGATGCTTGGGTCT 223
Qy 1375 gtccccaccacttcaggatgagatgcacacctccacacctcaggcccccttcagcactcc 1434
|||||
Db 224 GTTCCACCCACGTCAGGAGTGCCACCTCCACCTCAGGGCCCCCTTCAACCTGACTCC 283
Qy 1435 tccctaccgcagtttgagcaatacagaacggagagactttcactcagtttaccagctct 1494
|||||
Db 284 TCCCTTACCACCAATTTTGAACATCAACAGCAGGAGAGTGTGATCTGTTTATTCGCGCCT 343
Qy 1495 atcagtggtgcatcatcgcgaagcagggccagcacatcagcagcttctcgtttgc 1554
|||||
Db 344 GTCCCTTGCGCCATCATTTGGCAACAGCGGCCAACACATCAACAGCTTCTCGTTCG 403
Qy 1555 tggagcttcaattaaagattgtctccagcgggaagcaccagatgctaaagtggagtgat 1614
|||||
Db 404 GGGAGCTCGATTAAAGATCGCTCCAGCGGAGCAGCAGATGCTAAAGTGGCGATGGTAT 463
Qy 1615 tatcaatggaccaccagaggtcagttcaaggtcaggaagaatttatgaaaaataa 1674
|||||
Db 464 TATCACTGGACCACAGAGGTCAGTTCAGGCTCAGGGAAGAAATTTATGAAAAATTA 523
Qy 1675 aagaagaaactttgttagtctaaagagaggtgaaactgaagctcatatcagagtcc 1734
|||||
Db 524 AGAAGAACACTTTGTTAGTCTTAAAGAGAGGTGAACCTTGAAGCTCACATCAGAGTCC 583
Qy 1735 atcctttgctgctgcagaggtatttgaaaaaggaggcgaacagggtgaatgaacttcagaa 1794
|||||

Db 584 GTCTTTTGTCTGCTGCAGAGTATTGTTGGGAAAGGAGGCAAACTGTGAATGAGCTCCAGAG 643
Qy 1795 ttgtcaagtgcagaagttgtgtccctctgtagccagcacactgatgagaatgaccaagt 1854
|||||
Db 644 TTTATCACGTGCTTGAGTGTGCTCCCGCTGACCCAGACACCTGATGAGAATGATCAAT 703
Qy 1855 ggtgtcaaaaataactggtgcactctatgc-ttgcaggtgtcccgagagaaaaatcagg 1913
|||||
Db 704 AGTTGTCAACATAACTGGCCACTTCTATGCTTTGCGGGTCCCGAGAGAAATTCAGG 763
Qy 1914 aaattc 1919
|||||
Db 764 AAATTC 769
|||||
RESULT 15
BG499373 BG499373 633 bp mRNA EST 27-MAR-2001
LOCUS 602547471F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4670008 5',
DEFINITION mRNA sequence.
ACCESSION BG499373
VERSION BG499373.1 GI:13460890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 633)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-riemail.nih.gov
Tissue Procurement: DCTP/DNP
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CMI482 row: m column: 17
High quality sequence stop: 618.
Location/Qualifiers
1. .633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4670008"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: Sfil (ggcgctcgcc); Site_2: Sfil (ggcattatggcc
) ; Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCATATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 219 a 135 c 145 g 134 t
ORIGIN
Query Match 14.6%; Score 610.6; DB 11; Length 633;
Best Local Similarity 98.6%; Pred. No. 1.3e-87;
Matches 616; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1616 atcactggaccaccagaggtcagttcaggtcaggggaagaatttatgaaaaataa 1675
|||||
Db 1 ATCACTGGACCACAGAGGCTCAGTCAAGGCTCAGGGAAGAAATTTATGAAAAATTA 60

QY 1676 gaagaaaactttttagtcctctaagaagaggtgaacctgaagctcctatatcagagtgcca 1735
|||||
Db 61 GAAGAAAACCTTGTAGTCTCTAAAGAGAGGTGAAACTTGAAGCTCATATCAGAGTGCCA 120
QY 1736 tcctttgctgctgagcagagttattgaaagagaggaacaaacggtgaatgaacttcagaat 1795
|||||
Db 121 TCCTTTGCTGCTGCAGAGTTATTGGAAAAGGAGGCAAAACGGTGAATGAACTTCAAAAT 180
QY 1796 ttgtcaagtgacagaagtgtgttcctctgtaccagacacctgatgagaatgaccaagt 1855
|||||
Db 181 TTGTCAAGTGCAGAAAGTTGTGTCCCTCGTGACCAGACACCTGATGAGAAATGACCAAGTG 240
QY 1856 gttgtcaaaaataactgctactctctatgcttgcagggttcccagagaaaaattcaggaa 1915
|||||
Db 241 GTTGTCAAAAATAACTGCTCACTTCTATGCTTGCCAGGTTGCCAGAGAAAAAATTCAGGAA 300
QY 1916 attctactcaggtaaagcagaccacaaacacagaaggctctgcaaaagtgcaccacctcag 1975
|||||
Db 301 ATTCTGACTCAGGTAAAGCAGCACCAACAACAGAGGCTCTGCAAAAGTGGACCACCTCAG 360
QY 1976 tcaagcgggaagtaaaggctcaggaacacagccaccacagaggcagatgccaaaccaaag 2035
|||||
Db 361 TCAAGACGGAAGTAAAGGCTCAGGAACACAGCCACCACAGAGGCAGATGCCAAACCAAG 420
QY 2036 acagattgcttaaccaacagatggcgctgacccctctatccagaatcacatgcacaagtt 2095
|||||
Db 421 ACAGATTGCTTAAACCAACAGATGGCGCTGACCCCTATCCAGAATCACAATGCACAAGTT 480
QY 2096 tttaacctagccagttgttctcagagccaggaacttttgaactcctgtctctgtgagaa 2155
|||||
Db 481 TTTACCTAGCCAGTTGTTCTGAGGACCCAGGCAACTTTTGAACCTCCTGTCTGTGAGAA 540
QY 2156 tgtatactttatgctctctgtaaatgtatgacccccagcttttaaaacaaacaaacaaacaa 2215
|||||
Db 541 TGTATACTTTTATGCTCTCTGAATGTATGACACCCAGCTTTTAAACAACAACAACAACAA 600
QY 2216 acaaaaaaagggtgggggagggagg 2240
|||||
Db 601 ACAAAAAAAAAAGGGTGGGGCAGGG 625

OM of: US-09-685-696-176 to: N_Geneseq_1101.* out_format : pfs
Date: Jan 25, 2002 11:25 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09685696/runat_25012002_145551_10682/app_query.fasta_1.648
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -LOOPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=dylosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09685696_CGNI_1_331 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-685-696-176
Query length: 579
Database: N_Geneseq_1101.*
Database sequences: 930621
Database length: 428662619
Search time (sec): 163.950000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	2956.00	4242.42	4159	3.1e-228
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	2956.00	4242.37	4181	3.2e-228
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	2943.00	4232.43	1740	1.1e-227
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	1950.50	2792.34	2224	1.0e-166
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	1873.00	2686.08	3412	1.9e-147
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	1868.00	2673.59	1946	1.5e-141
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	1637.00	2346.58	3283	7.6e-141
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	702.50	1005.89	1708	1.3e-122
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	480.50	921.09	710	5.9e-48
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	440.50	886.47	444	3.1e-43
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	474.00	885.22	588	3.7e-30
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	341.00	490.57	282	4.3e-19
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	319.50	457.99	364	3.0e-19
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	284.50	410.93	424	2.0e-17
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	241.00	327.21	300	8.2e-15
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	231.50	329.24	500	3.8e-10
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	229.50	313.42	1420	2.9e-10
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	228.50	309.17	1803	1.1e-09
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	227.00	335.87	2381	2.2e-09
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	227.00	303.65	3319	3.8e-09
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	219.50	296.11	2396	2.0e-08
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	217.50	295.00	2007	2.3e-08
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	213.50	286.66	2588	6.8e-08
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	204.50	273.85	2541	3.5e-07
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	203.50	267.29	4219	8.2e-07
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	203.50	266.98	4353	8.6e-07
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	202.00	268.90	2901	6.7e-07
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	191.50	261.40	1357	1.7e-06
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	190.00	259.00	1389	2.4e-06
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	187.00	251.17	1392	2.4e-06
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	187.00	251.17	1966	6.5e-06
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	183.00	243.37	2401	1.8e-05
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	183.00	240.92	3061	2.4e-05
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	172.00	230.45	1790	9.3e-05
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	170.50	226.72	2090	0.0001
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	165.50	230.35	713	9.4e-05
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150	..	159.50	214.76	1416	0.0007

/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAH14076 + 158.00 208.68 0.0015
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV41260 + 157.50 204.83 0.0025
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ46281 + 154.50 204.47 0.0026
/SID88/gcgdata/geneseq/geneseq/NA1992.DAT:AAQ30801 + 154.00 200.11 0.0045
/SID88/gcgdata/geneseq/geneseq/NA1996.DAT:AAZ30564 + 153.50 204.96 0.0024

seq_name: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36150

seq_documentation_block:

ID AAZ36150 standard; DNA; 4159 BP.
AC AAZ36150;
DT 11-FEB-2000 (first entry)
DE DNA encoding cancer associated antigen KOC-1.
KW Cancer associated antigen; KOC-1; cancer; vaccine; CT7; ss.
OS Homo sapiens.
PN W09954738-Al.
XX 28-OCT-1999.
PF 16-MAR-1999; 99WO-US05766.
PR 17-APR-1998; 98US-0061709.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
DR WPI; 2000-013284/01.
XX Nucleotides representing cancer-associated genes, used to develop products for the diagnosis, monitoring and treatment of cancers -

Claim 88; Page 39-40; 44pp; English.
CC The present sequence represents a cancer associated antigen gene designated KOC-1. The specification also describes a cancer associated antigen designated CT7. The CT7 polypeptide was isolated from SK-MEL-37 melanoma cells. The polypeptide has some homology with MAGS-10, limited to about 210 carboxy terminal amino acids. The amino terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AA143877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony stimulating factor (GM-CSF).

Sequence 4159 BP; 1281 A; 830 C; 851 G; 1181 T; 16 other;

alignment_scores:
Quality: 2956.00 Length: 579
Ratio: 5.105 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAZ36150 ..
Align seg 1/1 to: AAZ36150 from: 1 to: 4159
1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
251 ATGAACAACAACTGTATATCGGAACCTCAGGAGAACGCCGCCCTCGGA 300
17 pLeuGluSerIlePheLysAspAlaTyrIleProValSerGlyProPheL 34

```
|||||
301 CCTAGAAAGTATCTTCAGGAGGCCAAGATCCCGGTGTCGGGACCCCTCC 350
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
351 TGGTGAAGACTGGCTACGCGTTCTGGGACTGCCCGGACGAGAGCTGGGCC 400
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
401 CTCGAAGGCCATCGAGGCCCTTTCAGGTAAATAGAACTGCACGGGAAACC 450
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
451 CATAGAAGTTGAGCACTCGGTCCCAAAAGGCAAGGATTTCGGAACCTC 500
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
501 AGATACGAAATATCCGCTCATTTACAGTGGGAGTCTGGATAGTTTA 550
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
551 CTAGTCCAAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
|||||
601 GGAAACTCAGTTGTAATGTAACTATTCCAGTAAGGACCAAGCTAGAC 650
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
651 AAGCACTAGACAAAGTGAATGGATTTACGTTAGAGAATTTACACCTTGAA 700
151 ValAlaTyrIleProAspGluMetAlaGlnGlnAsnProLeuGlnG 167
701 GTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGCAGCA 750
167 nProArgLysArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
751 GCCCGGAGTCCCGGGGGCTTGGCGAGAGGGGCTCCTCAAGGCGAGGGGT 800
184 erProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
801 CTCAGGATCCGTATCCAAGCAGAAACCATGTGATTTGCCTCTGGCCTG 850
201 LeuValProThrGlnPheValGlyAlaIleleGlyLysGluGlyAlaTh 217
851 CTGGTCTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGGTGCCAC 900
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
901 CATTCGGNACATACCAACAGACCCAGCTCTAAATCGATGCCACCGTA 950
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
951 AAGAAATGCGGGGCTGCTCAGAGAATCGATTACTATCTCTACTCCT 1000
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
1001 GAAGGCACCTCTCGCGGCTGTGAAGTCTATTCTGGAGATTATGCATAAGGA 1050
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
1051 AGCTCAAGATATAAATTCAGAGAAGAGATCCCTTGAAGATTFTAGCTC 1100
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
1101 ATAATACTTTGTTGGAGCTTATTGTTAAAGAGGAGGAGAAATCTTAA 1150
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
1151 AAAATTGAGCAAGACACAGACACTTAAATACGATATCTCATTTGCAGGA 1200
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
```

.. 4

```
1201 ATTGACGCTGTATATATCCAGAACGCACCTATTACAGTTAAAGCAATGTTG 1250
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1251 AGACATGTGCCAAAGCTGAGGAGGAGATCATGAAGAAATCAGGGAGTCT 1300
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
|||||
1301 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAAATTCCTGG 1350
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerGlyMetProp 384
|||||
1351 ATTAATCTGAACGCTTGGGTCTGTCCACCCACTTCAGGGATGCCAC 1400
384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1401 CTCACACCTCAGGGCCCTTTCAGCATGACTCTCCCTACCCGCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1451 GAGCAATCAGAAACGGAGACTGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleleGlyLysGlnGlyGlnHisLysGlnLeuSerArgP 434
|||||
1501 CGGTGCCATCATCGCAAGCAGGCCAGCATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1551 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAGCACCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG 467
|||||
1601 GTGAGGATGGTGATTATCCTGAGCAGCCAGAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1651 GGGAGGAATTTATGAAATAATTAAGAAAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1701 AAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTTTGCTGCTGGC 1750
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1751 AGAGTTATTGGAAGAGGAGGCAAAACGCGTAATGAACCTCAGAATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1801 AAGTGCAGAGTGTGTGCTCTGTCAGCAGACACCTGATGAGAAATGACC 1850
534 InValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1851 AAGTGGTTGCAAAATAACTGGTCACTTCTATGCTGCCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnLy 567
|||||
1901 AGAAAAATTCAGGAATTTCTGACTCAGTAAAGCAGCACCACCAACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1951 GGCTCTCAAAAGTGGACCACTCAGTCAAGACGGAAG 1987
seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC65900
seq_documentation_block:
ID AAC65900 standard; cDNA; 4181 BP.
XX
AC AAC65900;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated cDNA L523S;
XX
```

KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
 KW vaccine; detection; ss.

XX Homo sapiens.

XX WO200061612-A2.

XX 19-OCT-2000.

XX 03-APR-2000; 2000WO-US08896.

XX 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX (CORI-) CORIXA CORP.

PI Wang T, Fan L;

XX WPI; 2000-628399/60.

DR P-FSDB; AB111328.

XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
 PT protein is used for detecting and monitoring progression of lung cancer
 PT in a patient -

XX Claim 1a; Page 184-186; 261pp; English.

XX This invention describes a novel isolated polypeptide (I) which
 CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
 CC which have cytostatic activity. The polypeptides and polynucleotides are
 CC used in compositions and vaccines to inhibit the development of cancer,
 CC especially lung cancer, in a patient. Methods described in the invention
 CC can be used to monitor the progression of a cancer by carrying out the
 CC detection at subsequent time points and comparing the results from the
 CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
 CC are treated with P2, polynucleotides encoding P2 or antigen presenting
 CC cells expressing P2 and then administered to the patient to inhibit
 CC development of cancer.

XX Sequence 4181 BP; 1303 A; 830 C; 851 G; 1181 T; 16 other;

alignment_scores:

Quality: 2956.00 Length: 579
 Ratio: 5.105 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAC65900 ..

Align seg 1/1 to: AAC65900 from: 1 to: 4181

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17

251 ATGAACAACTGTATATCGAAACCTCAGGAGAACGCCGCCCTCGGA 300

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34

301 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGCGGGACCCCTTC 350

34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50

351 TGGTGAAGACTGGCTACCGGCTTCGTGGACTGCCCGGACGAGACTGGGCC 400

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67

401 CTCAGGCCATCGAGCGCTTCAGGTAATAATAGAACTGCACGGGAACC 450

67 olleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84

|||||

451 CATAGAAGTTGAGCACTCGGTCCCAAAAAAGGCAAGGATTTCGGAAACTTC 500
 84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
 |||||
 501 AGATACGAATATCCCGCCCTCATTTACAGTGGGAGGTGCTGGATAGTTTA 550
 101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
 |||||
 551 CTAGTCCAGTATGGAGTGGTGGAGAGCTGTGAGCAAGTGAACACTGACTC 600
 117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
 |||||
 601 GGAACCTGCAGTTGTAATGTAACCTATTCCAGTAAGGACCAAGCTAGAC 650
 134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
 |||||
 651 AAGCACTACACAAACTGAATGGATTTTCAGTTAGAGAAATTTCACTTGA 700
 151 ValAlaTyrIleProAspGluMetalAlaGlnGlnAsnProLeuGlnGl 167
 |||||
 701 GTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGACGCA 750
 167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
 |||||
 751 GCCCCAGAGTCGCCGGGGCTTGGGCAGAGGGCTCTCAAGCAGCGGGT 800
 184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
 |||||
 801 CTCAGGATCCGTATCCAGCAGAAACCATGCTGATTTGCCCTCTGCCGCTG 850
 201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
 |||||
 851 CTGGTTCCACCCCAATTTGTTGGAGCATCATAGGAAAGAGGTGCCAC 900
 217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
 |||||
 901 CAFTCCGAACATCACCACAAACAGACCAGTCTAAATCGATGTCCACCGTA 950
 234 ySgluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
 |||||
 951 AAGAAATCGGGGGCTGCTGAGAAGTCGATTACTATCTCTCTACTCCT 1000
 251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMethHisLysG 267
 |||||
 1001 GAAGGCACCTCTCGCGCTTGTAAAGTCTATTCTGGAGATTATGCATAAGA 1050
 267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
 |||||
 1051 AGCTCAAGATATAAAATTCACAGAAGAGATCCCTTTGAAGATTTTAGCTC 1100
 284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
 |||||
 1101 AFAATAACTTTGTTGGACGCTTATTGGTAAAGAGGAAAGAAATCTTAAA 1150
 301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnGl 317
 |||||
 1151 AAAATTTGAGCAAGACACAGACACTAAATCAGCATATCTCCATTGCAGGA 1200
 317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
 |||||
 1201 ATTGACGCTGTATAATCCAGAACGCACATATTACAGTTAAAGGCAATTTG 1250
 334 luThrCysAlaLysAlaGluGluIleMethLysLysIleArgGluSer 350
 |||||
 1251 AGACATGTGCCAAGCTTGAGGAGGAGATCATGAAGAAATCAGGGAGTCT 1300
 351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGl 367
 |||||
 1301 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATCTCTGG 1350
 367 yLeuAsnLeuAsnAlaLeuGlyIlePheProProThrSerGlyMetProp 384
 |||||
 1351 ATTAATCTGAACGCTTGGGTCTGTGCCACCCACTTCAGGGATGCCAC 1400

384 roProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1401 CTCACCTCAGGCGCCCTTCAGGCATGACTCCTCCTACCCGAGTTT 1450
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1451 GAGCAATCAGAAACGAGACATGTTTCATCAGTTTATCCAGCTCTATCAGT 1500
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgp 434
CGGTGCCATCATCGCAAGCAGGCGCAGACATCAAGCAGCTTCTCGCT 1550
434 heAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1551 TTGCTGGAGCTTCAATTAAAGATTTCTCCAGCGGAAGCACCAGATGCTAAA 1600
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG1 467
|||||
1601 GTGAGGATGGTATTCATCTGACCTGGACACAGAGGCTCAGTTCAAGGCTCA 1650
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1651 GCGAAGATTTATGAAAAATTAAAGAGAAACTTTGTTAGTCTCTAAAG 1700
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1701 AAGAGGTGAACCTTGAAGCTCATATCAGAGTGCCATCCTTTGCTGCTGGC 1750
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1751 AGAGTTATGGAAAGGAGGCAAAACGGTGAATGAACCTTCAGAATTTGTC 1800
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1801 AAGTGCAGAACTTCTGCTCCCTCGTACCAGACACCTGATGAGATGACC 1850
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1851 AAGTGGTGTGCAAAATAACTGGTCACCTCTATGCTTGCAGGTTGCCAG 1900
551 ArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGlnLy 567
|||||
1901 AGAAAAATTCAGGAATTCAGTACAGTAAAGCAGCACCACCAACACAGAA 1950
567 sAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
|||||
1951 GGCTCTGCAAGTGCACACCTCAGTCAAGAGCGGAAG 1987

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT: AAC66035

seq_documentation_block:

ID AAC66035 standard; cDNA; 1740 BP.

XX AAC66035;

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
XX Wang T, Fan L;
XX
XX WPI; 2000-628399/60.
XX P-PSDB; AAB11365.
XX
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX protein is used for detecting and monitoring progression of lung cancer
XX in a patient -
XX
XX Claim 1a; Page 258-259; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2 and then administered to the patient to inhibit
XX development of cancer.
XX
XX Sequence 1740 BP; 526 A; 406 C; 417 G; 391 T; 0 other;

alignment_scores:

Quality: 2943.00 Length: 579

Ratio: 5.101 Gaps: 0

Percent Similarity: 99.655 Percent Identity: 99.655

alignment_block:

US-09-685-696-176 x AAC66035 ..

Align seg 1/1 to: AAC66035 from: 1 to: 1740

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
|||||
1 ATGAACAACTGTATATCGGAAACCTCAGCGAGAACGCGCCCTCGGA 50
17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyPheL 34
|||||
51 CCTAGAAAGTATCTTCAAGGACGCCAAGATCCCGGTGTCGGGACCCCTTC 100
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
|||||
101 TGGTGAAGACTGGCTACGCTTCGTGGACTGCCCGGAGAGAGCTGGGCC 150
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
151 CTCAAGGCCATCGAGGCGCTTTCAGGTAAATAGAACTGCACGGGAACC 200
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
201 CATAGAAGTTGAGCAGCTCGGTCCCAAAAGCAAGGATTCGGAAACTTC 250
84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
|||||
251 AGATACGAAATATCCCGCCTCATTTACAGTGGGAGGTGCTGGATAGTTA 300
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
|||||
301 CTAGTCCAGTATGAGTGGTGGAGAGCTGTGACGACAGTGAACACTGACTC 350
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
|||||
351 GGAACATGCAGTTGTAANTGTACCTATTCCAGTAGGACCAAGCTAGAC 400

seq_documentation_block:

ID AAC66035 standard; cDNA; 1740 BP.

XX AAC66035;

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
|||||
401 AAGCACTAGACAACTGAATGATTTACGTTAGAGAATTTACCTTTGAAA 450
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167
|||||
451 GTAGGCTATATCCCTGATGAAGCGGCCGCCAGCAAAACCCCTTCAGCA 500
167 nProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
|||||
501 GCGCCGAGTCCCGGGGGCTTGGCGAGAGGGCTCTCTCAAGCGAGGGT 550
184 exProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeu 200
|||||
551 CTCCAGGATCCGTATCCAGCAGCAAAACCATGTGATTTGCTCTCGCGCTG 600
201 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaTh 217
|||||
601 CTGGTTCCACCCCAATTTGTTGGAGCCATCATAGGAAAGAGAGTGCCAC 650
217 rIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArgL 234
|||||
651 CATTCGGAACATCACCAACAGACCCAGTCTAAATCGATGTCACCGTA 700
234 ysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThrPro 250
|||||
701 AAGAAATCGGGGGCTGCTGAGAAGTCGATTACTCTCTCTACTCTCT 750
251 GluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysG 267
|||||
751 GAAGGCACCTCTCGCGCTTGTAAAGTCTATTCTGGAGATTATGCATAAG 800
267 uAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaH 284
|||||
801 AGCTCAAGATATAAAATTCAGAGAGATCCCTTGAAGATTTTAGCTC 850
284 isAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 300
|||||
851 ATAATAACTTTGTTGGACGCTCTATTGGTAAAGAGAAATCTTAA 900
301 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnG 317
|||||
901 AAAATTGAGCAGACACAGACATAAAATCAGATATCTCCATTGCAGGA 950
317 uLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnValG 334
|||||
951 ATTGACGCTGATAATCCAGAACGCACTATTACAGTTAAAGCAATGTTG 1000
334 luThrCysAlaLysAlaGluGluIleMetLysLysIleArgGluSer 350
|||||
1001 AGACATGTCGCAAGCTGAGGAGGAGATCATGAGAAATCAGGGAGTCT 1050
351 TyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProG 367
|||||
1051 TATGAAATGATATTGCTTCTATGAATCTTCAAGCACATTTAATTCCTG 1100
367 yLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProp 384
|||||
1101 ATTAATCTGAACGCCCTGGGTCTGTGCCACCCCACTTCAGGGATGCCAC 1150
384 rProThrSerGlyProProSerAlaMetThrProProTyrProGlnPhe 400
|||||
1151 CTCACCACTCAGGGCCCCCTTACGCCATCAGCTCTCTCCCTACCCGCACTT 1200
401 GluGlnSerGluThrGluThrValHisGlnPheIleProAlaLeuSerVa 417
|||||
1201 GAGCAATCAGAAAGGAGAGCTGTTCTATCTCTTTATCCAGCTCTATCAGT 1250
417 lGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgP 434
|||||
1251 CGGTGCCATCATCGGCAAGCAGGGCCAGCACATCAAGCAGCTTCTCGCT 1300

434 heaGlyAlaSerIleLysIleAlaProAlaGluAlaProAspAlaLys 450
|||||
1301 TTGCTGGAGCTTCAATTAAGATTGCTCCAGCGGAAGCACCAGATGCTAAA 1350
451 ValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAlaG 467
|||||
1351 GTGAGGATGGTGATTATCACTGGACCACCAGAGGCTCAGTTCAAGGCTCA 1400
467 nGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysG 484
|||||
1401 GCGAACAATTTATGGAAAAATTAAGAAGAAAACTTTGTAGTCTCTAAG 1450
484 luGluValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGly 500
|||||
1451 AAGAGGTGAAACTTGAAGCTCATATCAGAGTGCATCTCTTTGCTGTGCG 1500
501 ArgValIleGlyLysGlyLysThrValAsnGluLeuGlnAsnLeuSe 517
|||||
1501 AGAGTTATTGGAAAAGGAGGCAAAACGGTGAATCAACTTCAGAAATTTGTC 1550
517 rSerAlaGluValValProArgAspGlnThrProAspGluAsnAspG 534
|||||
1551 AAGTGAGAAAGTTGTTGCTCTGTCGACACACACCTGATGAGATGACC 1600
534 lnValValValLysIleThrGlyHisPheTyrAlaCysGlnValAlaGln 550
|||||
1601 AAGTGGTTGTCAAAATAACTGGTCACTTCTATGCTTGCCAGGTGCCCAG 1650
551 ArgLysIleGlnGluLeuThrGlnValLysGlnHisGlnGlnLys 567
|||||
1651 AGAAAAATTCAGGAAATTCGACTCAGGTAAGCAGCACCACCAACAGAA 1700
567 sAlaLeuGlnSerGlyProGlnSerArgArgLys 579
|||||
1701 GGCTCTGCAAAAGTGGACCACCTCAGTCAAGACCGGAAG 1737

seq_name: /SID88/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ10617

seq_documentation_block:

ID AAZ10617 standard; cDNA; 2224 BP.

XX AAZ10617;

XX 17-NOV-1999 (first entry)

XX cDNA encoding a murine c-myc coding region determinant binding protein.

XX c-myc coding region determinant binding protein; CRD-BP; tumor;
KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer; ss.

OS Mus musculus.

XX Key Location/Qualifiers

XX CDS 131..1854

XX FT /*tag= a

XX PN WO9946594-A2.

XX PD 16-SEP-1999.

XX PF 05-MAR-1999; 99WO-US04897.

XX PR 09-MAR-1998; 98US-0077372.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX PI Ross J;

XX WPI; 1999-551506/46.

DR P-PSDB; AAY30649.

XX Diagnosing presence or absence of a tumor in a human by examining c-myc

PT coding region determinant-binding protein

XX Example; Fig 1A-D; 79pp; English.

XX The present sequence encodes a murine c-myc coding region determinant binding protein (CRD-BP). The presence or absence of a tumor can be determined by determining the levels of CRD-BP present in the suspect tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack and so prolongs its half-life. The methods are used for diagnosing CC presence or absence of a tumor in a human, especially breast, colon and pancreatic cancer. They are also used to inhibit cancer cell growth.

XX Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 other;

alignment_scores:

Quality: 2187.00 Length: 585
Ratio: 4.150 Gaps: 8

Percent Similarity: 90.085 Percent Identity: 74.188

alignment_block:

US-09-685-696-176 x AAZ10617 ..

Align seg 1/1 to: AAZ10617 from: 1 to: 2224

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaProSerAs 17
131 ATGAACAAGCTTTACATCGGCAACCTCAACGAGAGTGTGACCCCGCAGA 180
17 pLeuGluSerIlePheIysAspAlaLysIleProValSerGlyProPheL 34
181 CTGGAGAAAGTATTCTCGGAGCACAAGATCTCTACAGCGCGCGAGTTCT 230
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
231 TGGTCAATCGGGTACGCTTCGTGGATGCCCCGACGACACTGGGCG 280
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
281 ATGAAGGCCATCGAAACTTTCTCGGGGAAAGTAGAACTGCAAGGAAACG 330
67 oIleGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
331 TCTAGAGATTGAACACTCAGTCCCAAAACAAAGAGGAGTCGGAATAAC 380
84 InIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
381 AGATCCCGCAATATTCACCTCAGCTCCGATGGGAAGTGTAGATAGCTG 430
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
431 CTGGCTCAGTACGGTACAGTGGAGAACTGTGAGCAAGTGAACACTGAAAG 480
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArg 134
481 TGAGACAGCGGTGGTCAACGCTCACTACTTAACCGGGAGCAGACCAAGC 530
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
531 AAGCTATCATCAAGCTAAATGGCCATCACTGGGAACCATGCCCTGAAG 580
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
581 GTCTCTACATACCTGATGACGAGATAACA...CAAGGTCTCTGAGAATGG 627
167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGlyS 184
628 GCGTCGTGGA.....GGCTTTGGGTCTCGGGGCCAGCCCGCAAGGGT 671
184 erPro.....GlySerValSerGlnLysProCysAspLeuPro 197
672 CGCCCGTGGCAGCGGGCTCCAGCCCAAGCAGCAGCAGTGGACATCCCT 721

198 LeuArgLeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysG 214
722 CTCGGGCTCCTGGTGGCTACGCAAGTATGTAGCGCTATCATTTGCAAGGA 771
214 uGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspV 231
772 GGGTGGCCACCATCGCAACATCACAACACAGCAGCTCCAAATAGACG 821
231 alHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeu 247
822 TGCATAGGAAGAGATCGGGCGCTCGGAGAGAGGCATCAGCGTGCAT 871
248 SerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMe 264
872 TCAACCCCTGAAGGCTGCTCCTCGCGTGCAGATGATCTTTGGAGATTAT 921
264 tHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysI 281
922 GCACAAGGAGGCAAGGACACCAACACGGCAGATGAAGTTCCCTGAAGA 971
281 leLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArg 297
972 TCCTGGCTCATAACAACCTTCGTGGGCGACTCATTTGGCAAGAGGCGCG 1021
298 AsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleSerPr 314
1022 AACCTGAAGAAGGTGGAGCAGCAGACAGAGCAAGATCACCATCTCATC 1071
314 oLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysG 331
1072 GCTCCAGGACCTCACGCTCTATAACCTCGAGAGGACCATCATCTGAAG 1121
331 lyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIle 347
1122 CGGCATTGGAACCTGTTGAGGCGCCGAGCAGGAGATCATGAAGAAGTT 1171
348 ArgGluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLe 364
1172 CGAGAGGCTTACGAGAACGAGCTGGCGCCCATGAGTTGCAGTCCACCT 1221
364 uIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProThrSerG 381
1222 CATCCCTGGGCTTAACTGGCTGCTGAGGTCTCTCCAGCTTCCATCCA 1271
381 lyMetProProThrSerGlyProProSerAlaMetThr.....Pro 395
1272 GCGTGTCTCCTCT.....CCTCCAGCAGTGTCACTGGGGCTGCT 1312
396 ProTyrProGlnPheGluGlnSer...GluThrGluThrValHisGlnPh 411
1313 CCCTATAGTCTCTCATGCAAGCTCCGAGCAGGAGATGGTACAAGTGT 1362
411 eIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlnHisI 428
1363 CATCCCGCCCGAGGCTGGGCGCCATTTGGCAAGAGGCGCCAGCACA 1412
428 leLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 444
1413 TCAACAACCTCTCCCGTTTCGCCAGCGCTCCATCAAGATTGCTCCACCA 1462
445 GluAlaProAspAlaLysValArgMetValIleIleThrGlyProProGl 461
1463 GAAACACCTGACTCCAAAGTTCGAATGGTCTCATCTACTGGACCCCCAGA 1512
461 uAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLysGluAla 478
1513 GGCTCAGTTCAAGGCTCAGGGAAGATTTATGGCAAACTAAAGAGAGA 1562
478 snPheValSerProLysGluGluValLysLeuGluAlaHisIleArgVal 494
1563 ATTTCTTTGGTCCCAAGGAGAGTAGAAGCTAGAGACCCACATACGGGTT 1612

```
495 ProSerPheAlaAlaGlyArgValIleGlyLysGlyLysThrValAs 511
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1613 CCGGCTTCAGCAGCGCGCGCTCATCGGCAAGCGGCAAAACGGTGA 1662

511 nGluLeuGlnAsnLeuSerAlaGluValValProArgAspGlnT 528
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1663 TGAGCTGCAGAACTTACGTGCACCTGAGGTGGTAGTGCACAGACAGA 1712

528 hrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyr 544
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1713 CCGCGGATGAGAACGACCAAGTCAATGTTAAGATCATCGGACATTTCTAT 1762

545 AlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnVally 561
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1763 GCCAGCAGATGGCTCAGCGAAGATCCGAGACATCTGGCTCAAGTTAA 1812

561 sGlnHisGlnGlnLysAlaLeuGlnSerGlyProProGlnSerArgA 578
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1813 GCAA...CAGCACCAGAAGGA...CAGAGCAACCTGGCCCGCACGGCA 1856

578 rGlys 579
|||||
1857 GGAAG 1861
```

seq_name: /SID88/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ36152

seq_documentation_block:

ID AAZ36152 standard; DNA; 3412 BP.

XX AC AAZ36152;

XX DT 11-FEB-2000 (first entry)

XX DE DNA encoding cancer associated antigen KOC-3.

XX KW Cancer associated antigen; KOC-3; cancer; vaccine; CT7; ss.

XX OS Homo sapiens.

XX PN W0954738-Al.

XX PD 28-OCT-1999.

XX PF 16-MAR-1999; 99WO-US05766.

XX PR 17-APR-1998; 98US-0061709.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX DR WPI; 2000-013284/01.

XX PT Nucleotides representing cancer-associated genes, used to develop

XX products for the diagnosis, monitoring and treatment of cancers

XX PS Claim 55; Page 41; 44pp; English.

XX CC The present sequence represents a cancer associated antigen gene
CC designated KOC-3. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polynucleotide was isolated from
CC SK-MEL-37 melanoma cells. The polypeptide has some homology with
CC MAGE-10, limited to about 210 carboxy terminal amino acids. The amino
CC terminal of the protein has a repetitive pattern, with repeats rich in
CC serine, proline, glutamine and leucine, and an almost invariable core of
CC the peptide given in AA43877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony
CC stimulating factor (GM-CSF).

```
SQ Sequence 3412 BP; 970 A; 887 C; 804 G; 750 T; 1 other;

alignment_scores:
  Quality: 1950.50      Length: 591
  Ratio: 3.817         Gaps: 8
  Percent Similarity: 86.464  Percent Identity: 65.313

alignment_block:
US-09-685-696-176 x AAZ36152 ..

Align seg 1/1 to: AAZ36152 from: 1 to: 3412

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaLaprosSerAs 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
73 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGCTCACCGCGCAGCA 122

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
123 CTTCCGCGAGCTCTTTGGGACAGAGAAAGCTGCCCTGGCGGACAGGTCC 172

34 euValLysThrGlyTyrAlapheValAspCysProAspGluSerTrpAla 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
173 TGTGAAGTCCGGCTACGCTTCGTGGACTACCCCGACCAAGAACTGGGCC 222

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
223 ATCCGCGCATCGAGACCTCTCGGTTAAAGTGAATTGCATGGGAAAT 272

67 olleGluValGluHisSerValProLysArgGlnArgIleAargLysLeuG 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
273 CATGGAAGTTGATTACTACGTCTCTAAAAAGCTAAAGGAGCAGGAAATTC 322

84 lnIleArgAsnIleProProHisLeuGlnTrpGluValLeuAspSerLeu 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
323 AGATTGGAACATCCCTCCCTCAGCTGCAGTGGGAGGTGTTGGATGGACTT 372

101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
373 TTGGCTCAATATGGGACAGTGGAGAATGTGGACAAAGTCAACACAGACAC 422

117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
423 AGAAACCGCGCGTTGTCAACGTCACATATGCAACAAGAGAAGAAAGAAAA 472

134 lnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
473 TAGCCATGGAGAAGCTAAGCGGCATCAGTTTGAGAACTACTCCTTCAAG 522

151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGl 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
523 ATTTCTACATCCCGGATGAAGAGGTGAGTCCCTTCGCCCTCAG.. 570

167 nProArgGlyArgArgGlyLeuGlyGlnArgGlySerSerArgGlnGly. 183
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
571 ....CGAGCCAGCGT.....GGGGACCACCTCTTCCCGGAGCAGAGGCC 610

184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
611 ACGCCCTTGGGGCACTTCTCAGGCCACAGACAGATTGATTTCCTCCGCTCGG 660

200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
661 ATCCTGGTCCCAACCCAGTTGTTGGTGCCATCATCGGAAGAGAGGCGCTT 710

216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
711 GACCAATAAGAACATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATA 760

233 rgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
761 GAAAAGAGAACTCTGGAGCTGCAGAGAAGCGCTGTCAACCATCCATGCCACC 810
```

250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisIy 266
811 CCAGAGGGGACTTCGAAGCATGCCGCGATGATCTTGAATCATCGAGAA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
861 AGAGGCAGATGAGACCAAACTAGCCGAAGAGATTCTCTGAAATCTTGG 910
283 IaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
911 CACACAATGGCTTGGTTGGAAGACTGATTGGAAAAAGAGGAGCAAAATTG 960
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG1 316
961 AAAAAATGAACATGAACAGGACCAAGATAACAATCTCATCTTTGCA 1010
316 nGluLeuThrLeuYrAsnProGluArgThrIleThrValLysGlyAsnV 333
1011 GGATTGAGCATATACACCCGGAAGAACCATCACTCTGGAAGGCACAG 1060
333 aGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGlu 349
1061 TTGAGCGCTGTGCCAGTCTGAGATAGAGATTATGAAGAAGCTGCGTGAG 1110
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1111 GCCTTTGAAAATGATATCTGGCTGTTAAACCAACAAGCCAATCTGATCCC 1160
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPhe..... 376
1161 AGGTTGAACTTCACCGCAGCTTGGCATCTTTTCAACAGGAGTGTCCTGCC 1210
377ProProThrSer.....GlyMetProProProThr..... 386
1211 TATCTCCACAGCAGCGCCCGCGGAGCTCCCGCGCTGCCCTTACCAC 1260
387SerGlyProProSerAlaMetThrPro..... 395
1261 CCCTTCACTACCCACTCCGGATCACTTCCTCCAGCTGTATCCCTCCATCA 1310
396ProTyrProGlnPheGluGln...SerGluThrGluThrValH 409
1311 GTTTGGCCCGTTCGCGCATCACTCTTATCCAGCAGCAGGATTGTGA 1360
409 IsGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425
1361 ATCTCTTCATCCCAACCCAGCGCTGTGGCGCCATCATCGGAAGAAGGG 1410
426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAl 442
1411 GCACACATCAACAGCTGGCGAGATTCCCGGAGCCTCTATCAAGATTGC 1460
442 aProAlaGluAlaProAspAlaLysValArgMetValIleIleThrGlyP 459
1461 CCCTGCGGAAGCCAGACGTCAGGGAAGGATGGTCATCATCACCGGCG 1510
459 roProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysIleLys 475
1511 CACCGGAAGCCAGTTCAAGCCCGCAGGACGATCTTTGGGAAACTGAAA 1560
476 GluGluAsnPheValSerProLysGluGluValLysLeuGluAlaHisI1 492
1561 GAGGAAAATCTCTTAAACCCCAAGAGAGTGAAGCTGGAAGCGCATAT 1610
492 eArgValProSerPheAlaAlaGlyArgValIleGlyLysGlyLysT 509
1611 CAGAGTGCCTCTTCCACAGCTGCCGGGTGATTGGCAAAAGTGGCAGA 1660
509 hrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValProArg 525
1661 CCGTGAAACAACTGCAGAACTTAACAGTGCAGAAAGTCACTGCTGCTG 1710
526 AspGlnThrProAspGluAsnAspGlnValValLysIleThrGlyHI 542

|||||
1711 GACCAACGCCAGATGAAAATGAGGAAGTATCGTCAGAAATTATCGGCA 1760
542 sPheTyrAlaCysGlnValAlaGlnArgLysIleGlnGluIleLeuThrG 559
1761 CTTCTTTGCTACCGAGACTGCACACGCGAAGATCAGGGAATTTGTACAAC 1810
559 InValLysGlnHisGlnGlnGln 1866
1811 AGGTGAAGCAGCAGGAGCAAAA 1833

seq_name: /SID58/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ36153

seq_documentation_block:

ID AAZ36153 standard; DNA; 1946 BP.

XX AAZ36153;

XX 11-FEB-2000 (first entry)

XX An alternative form of DNA encoding cancer associated antigen KOC-2.

XX Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.

XX Homo sapiens.

XX WO9954738-A1.

XX 28-OCT-1999.

XX 16-MAR-1999; 99WO-US05766.

XX 17-APR-1998; 98US-0061709.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

XX WPI; 2000-013284/01.

XX Nucleotides representing cancer-associated genes, used to develop products for the diagnosis, monitoring and treatment of cancers -

XX Claim 55; Page 42; 44pp; English.

XX The present sequence represents an alternative form of a cancer associated antigen gene designated KOC-2. The specification also describes a cancer associated antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-37 melanoma cells. The polypeptide has some homology with MAG-10, limited to about 210 carboxy terminal amino acids. The amino terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AAY43877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or CC granulocyte macrophage-colony stimulating factor (GM-CSF).

XX Sequence 1946 BP; 502 A; 528 C; 553 G; 358 T; 5 other;

alignment_scores:

Quality: 1873.00 Length: 509

Ratio: 4.116 Gaps: 9

Percent Similarity: 89.391 Percent Identity: 74.656

alignment_block:

US-09-685-696-176 x AAZ36153 ..

Align seg 1/1 to: AAZ36153 from: 1 to: 1946

XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX WPI; 2000-013284/01.
XX
XX Nucleotides representing cancer-associated genes, used to develop
XX products for the diagnosis, monitoring and treatment of cancers -
XX
XX Claim 55; Page 42-43; 44pp; English.
XX
XX The present sequence represents an alternative form of a cancer
XX associated antigen gene designated KOC-3. The specification also
XX describes a cancer associated antigen designated CT7. The CT7
XX polynucleotide was isolated from SK-MEL-37 melanoma cells. The
XX polypeptide has some homology with MAGE-10, limited to about 210 carboxy
XX terminal amino acids. The amino terminal of the protein has a repetitive
XX pattern, with repeats rich in serine, proline, glutamine and leucine,
XX and an almost invariable core of the peptide given in AAY43877. The CT7
XX polypeptide can be processed to peptides which provoke lysis by
XX cytolytic T cells. The polynucleotides and polypeptides can be used for
XX treating a cancerous condition and screening for or diagnosing cancerous
XX conditions. The cancer associated antigens can be used as an immunogenic
XX or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or
XX granulocyte macrophage-colony stimulating factor (GM-CSF).
XX
XX Sequence 3283 BP; 945 A; 833 C; 779 G; 725 T; 1 other;

alignment_scores:
Quality: 1868.00 Length: 568
Ratio: 3.820 Gaps: 6
Percent Similarity: 86.092 Percent Identity: 64.789

alignment_block:
US-09-685-696-176 x AAZ36154 ..
Align seg 1/1 to: AAZ36154 from: 1 to: 3283

1 MetAsnLysLeuTyrIleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
73 ATGAACAAGCTTTATACATCGGAACCTGAGCCCGCCGCTCACCGCCGACGA 122
17 pLeuGluSerIlePheLysAspAlaIleProValSerGlyProPheL 34
123 CCTCCGGCAGCTCTTTGGGACAGGAAGCTGCCCTGGCGGACAGGCTC 172
34 euValLysThrGlyTyrAlaPheValAspCysProAspGluSerTrpAla 50
173 TGTGGAAGTCCGGCTACGCCCTTCGTGGACTACCCCGACAGAACTGGGCC 222
51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
223 ATCCGCCCATCGAGACCCCTCTCGGTAAAGTGAATGCGATGGGAAAT 272
67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
273 CATGGGAAGTTGATTACTACGTCTCTAAAAAGCTAAAGGAGCAGGAAATTC 322
84 InIleArgAsnIleProProHisGlnTrpGluValLeuAspSerLeu 100
323 AGATTGAAACATCCCTCCCTACCTCGAGTGGGAGGTTGGATGGACTT 372
101 LeuValGlnTyrGlyValValGluSerCysGluGlnValAsnThrAspSe 117
373 TTGGCTCAATATGGGACAGTGGAGATGTGGAACAAGTCAACACAGACAC 422
117 rGluThrAlaValValAsnValThrTyrSerSerLysAspGlnAlaArgG 134
423 AGAAACCGCGGTGTCAACGCTCATATGCAACAAGAGAGAGAGAGAAAA 472
134 InAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLys 150

473 TAGCCATGGAGAAGCTAAGCGGCATCAGTTTGAGAACTACTCTCTCAAG 522
151 ValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG1 167
523 ATTTCTTACATCCCGATGAAGAGGTGAGCTCCCTTCGCCCCCTCAG... 570
167 nProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGly. 183
571 ...CGAGCCAGGCT...GGGACCACTCTTCCCGGAGCAAGGCC 610
184 ..SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArg 199
611 AGCGCCCTTGGGGCACTTCTCAGCCAGACAGATTTCCCGCTCGGG 660
200 LeuLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAl 216
661 ATCTCGTGGTCCCAACCATGTTTGGTCCCATCATCGGAAAGAGGGCTT 710
216 aThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisA 233
711 GACCATAAAGAACATCACTAAGCAGACCCAGTCCCGGTAGATATCCATA 760
233 rGlyGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249
761 GAAAGAGAACTCTGGAGCTGCAGAGAGCTGTCCATCATCCATGCCACC 810
250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLy 266
811 CCAGAGGGGACTTCTGAAGCATCGGCATGATTCTTGAATCATGCAGAA 860
266 sGluAlaGlnAspIleLysPheThrGluGluIleProLeuLysIleLeuA 283
861 AGGCGAGATGAGACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGG 910
283 laHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 299
911 CACAAATGGCTTGGTGGAGACTGATTGGAAGAGAGGAGGAGAAATTTG 960
300 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuG1 316
961 AAGAAATTTGACATGAACAGGAGGACCAAGATAACATCTCATCTTTGCA 1010
316 nGluLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAsnV 333
1011 GGATTTGAGCATATACACCCGGAAGAACCATCACCTGTGAAGGCACAG 1060
333 alGluThrCysAlaLysAlaGluGluIleMetLysLysIleArgGlu 349
1061 TTGAGGCTGTGCCAGTCTGAGATGAGATTATGAAGAAGCTCGCTGAG 1110
350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIlePr 366
1111 GCCTTTGAAATGATATGCTGGCTGTTAAACCCACTCC... 1149
366 oGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetP 383
1150GGATACTTC..... 1158
383 roProProThrSerGlyProProSerAlaMetThrProProTyrProGln 399
1159TCCAGCTGTACCCCATCACCAGTTTGGCCGCTTCCCGCAT 1200
400 PheGluGln...SerGluThrGluThrValHisGlnPheIleProAlaLe 415
1201 CATCACTCTTATCCAGAGCAGGAGATGTTGAATCTCTTCATCCCAACCCA 1250
415 uSerValGlyAlaIleIleGlyLysGlnGlnHisIleLysGlnLeuS 432
1251 GGCTGTGGCGCCATCATCGGAAGAAGGGGACACATCAACAGAGCTGG 1300
432 eArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAlaProAsp 448
.....

```
1301 CGAGATTCGCCGAGCCTCTATCAAGATTGCCCTCGCGAAGCCCGAC 1350
449 AlaLysValArgMetValIleThrGlyProGluAlaGlnPheLeu 465
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1351 GTCAGCGAAGGATGGTATCATCATCCGGCCACCGGAAGCCAGTTCAA 1400
465 salAGlnGlyArgIleTyrGlyLysIleLysGluGluAsnProLeu 482
1401 GGCCAGGACGGATCTTTGGGAACTGAAGAGGAAACTCTTTAAACC 1450
482 rGLysGluValLysLeuGluAlaHisIleArgValProSerPheAla 498
1451 CCAAGAAGAGTGAAGCTGGAAGCGCATATCAGAGTGCCTCTCCACA 1500
499 AlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln 515
1501 GCTGGCGGGTATGGCAAGAGTGGCAAGACCGTGAAGCACTGCAGAA 1550
515 nLeuSerSerAlaGluValValProArgAspGlnThrProAspGluA 532
1551 CTTAACCACTGCAGAACTCATCGCTCGTGACCAAGCCAGATGAAA 1600
532 snAspGlnValValLysIleThrGlyHisPheTyrAlaCysGlnVal 548
1601 ATGAGGAAGTGATCGTCAAGATTATCGGGCACTCTTTGCTAGCCAG 1650
549 AlaGlnArgLysIleGlnGluLeuThrGlnValLysGlnHisGlnG 565
1651 GCACAGCGCAGATCAGGGAATTTGACAAAGGTGAAGCAGCAGGAC 1700
565 nGln 566
1701 GAAA 1704
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ36151

seq_documentation_block:

ID AAZ36151 standard; DNA; 1708 BP.

AC AAZ36151;

DT 11-FEB-2000 (first entry)

DE DNA encoding cancer associated antigen KOC-2.

KW Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.

OS Homo sapiens.

PN WO9954738-A1.

PD 28-OCT-1999.

PF 16-MAR-1999; 99WO-US05766.

PR 17-APR-1998; 98US-0061709.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;

DR WPI; 2000-013284/01.

PT Nucleotides representing cancer-associated genes, used to develop products for the diagnosis, monitoring and treatment of cancers -

PS Claim 55; Page 40; 44pp; English.

CC The present sequence represents a cancer associated antigen gene designated KOC-2. The specification also describes a cancer associated antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-37 melanoma cells. The polypeptide has some homology with MAGE-10, limited to about 210 carboxy terminal amino acids. The amino

CC terminal of the protein has a repetitive pattern, with repeats rich in serine, proline, glutamine and leucine, and an almost invariable core of the peptide given in AAV43877. The CT7 polypeptide can be processed to peptides which provoke lysis by cytolytic T cells. The polynucleotides and polypeptides can be used for treating a cancerous condition and screening for or diagnosing cancerous conditions. The cancer associated antigens can be used as an immunogenic or vaccine composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony stimulating factor (GM-CSF).

XX Sequence 1708 BP; 447 A; 469 C; 473 G; 314 T; 5 other;

alignment_scores:

Quality: 1637.00 Length: 453

Ratio: 4.093 Gaps: 9

Percent Similarity: 88.300 Percent Identity: 74.172

alignment_block:

US-09-685-696-176 x AAZ36151 ..

Align seg 1/1 to: AAZ36151 from: 1 to: 1708

```
133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLe 149
      ||| ||::: ||||| ||||| ||||| ||||| ||||| |||||
32 CGGGGAGCCATCATGAAGCTGAATGCCACCAGTTGGAGAACCATGCCCT 81
149 uLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuG 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 GAAGTCTCTTACATCCCGATGAGCAGATAGCA...CAGGGACCTGAGA 138
166 InGlnProArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGln 182
      :: ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 ATGGGCGCCGAGGG.....GGCTTTGGCTCTCGGGGTGAGCCGCCAG 172
183 GlySerPro.....GlySerValSerLysGlnLysProCysAspLe 196
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 GGCTCACCTGTGGCAGCGGGGCCCGCCAGCAGCAGCAAGTGGACAT 222
196 uProLeuArgLeuValProThrGlnPheValGlyAlaIleGlyL 213
      :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 CCCCCCTCGGCTCCTGGTCCCCACCCAGTATGTGGGTGCATTTATTGGCA 272
213 ySGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIle 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 AGGAGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAGATA 322
230 AspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrI 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 GACGTGCATAGGAAGGAGAACGCGAGCTGCAGCTGAAAAGCCATCAGTG 372
246 eLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluI 263
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 GCACCTCCACCCCTCGAGGGCTGCTCCTCCGCTTGAAGATGATCTTGAGA 422
263 leMethHisLysGluAlaGlnAspIleLysPheThrGluGluIleProLeu 279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 TTATGCTAAAGAGGCTAAGGACACCAACAGGCTGACGAGGTCCCTCG 472
280 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluI 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 AAGATCCTGGCCCATATAACTTTGTAGGCGCTCTCATTTGGCAAGGAAG 522
296 yArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThrIleS 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 ACGGAACCTGAGAGAGGTAGCAAGATACCGACAGACAAAAATCACCATCT 572
313 exProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrVal 329
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 CCTCGTGTGAAGACCTTACCCTTTACACCCCTGAGAGGACCATCAGTGT 622
330 LysGlyAsnValGluThrCysAlaLysAlaGluGluIleMetLysLy 346
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
187 erValSerLysGlnLysProCysAspLeuProLeuArgLeuLeuValPro 203
    ::::::::::: ::::::::::: |||||::|::|::|::|::|::|::|
114 GCACCTTCTCAGGCCAGACAGATGATTTCCCGCTGCGGATCTCGTCCCC 163
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
204 ThrGlnPheValGlyAlaIleuLeuGlyLysGluGlyAlaThrIleArgAs 220
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
164 ACCCATTTGTTGGTGCCCATCATCGGAAGAGGGGCTTGACCCATAAGAA 213
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
220 nIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnA 237
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
214 CATCACTAAGCAGACACCCAGCTCCCGGTAGATATCCATAGAAAAGAACT 263
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
237 laGlyAlaIaIaGluLysSerIleThrIleLeuSerThrProGluGlyThr 253
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
264 CTGGAGCTCGAGAAAGCCCTGTCCACATCATCATCCACCCAGAGGGACT 313
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
254 SerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAlaGlnAs 270
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
314 TCTGAAGCATGCCGATGATCTTTGAAATCATGCAGAAAGGACGATGA 363
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
270 pIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsp 287
    : ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
364 GACCAAACTAGCCGAAGAGATTCCTCTGAAAATCTTGGCACACAATGGCT 413
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
287 heValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLysIleGlu 303
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
414 TGGTTGGAAGACTGATTTGGAAGAAAGGAGCAAAATTTGAAGAAATTTGA 463
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
304 GlnAspThrAspThrLysIleThrIleSerProLeuGlnGluLeuThrLe 320
    ::::::::::: |||||::|::|::|::|::|::|::|::|::|::|::|
464 CATGAACAGGACCAAGATAACAACTCTCATCTTGCAGGATTTGAGCAT 513
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
320 uTyRAsnProGluArgThrIleThrValLysGlyAsnValGluThrCysA 337
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
514 ATACAACCCGGAAGAACCATCACTCTGAAGGCACAGTTGAGGCCCTGTG 563
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
337 laLysIaIaGluGluIleMetLysLysIleArgGluSerTyRAsn 353
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
564 CCAGTCTGAGATAGAGATTTATGAAGAACTGCGTGAGGCCTTTGAAAT 613
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
354 AspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGly...LeuAs 369
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
614 GATATCTGCTGTAACTGTAAGTCCCTAAATGCTTTCTTCNCGCTGGG 663
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
369 nLeuAsnAlaLeuGlyLeuPheProThrSer...GlyMetPro 383
    ::::::::::: ||| ::::::::::: ::::::::::: |||||::|
664 TTTCACTAGGCTAAAAATCTTGCCATTCAGCTNATGAGGAATGCCT 709
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

seq_name: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC17226

seq_documentation_block:

ID AAC17226 standard; cDNA; 444 BP.

AC AAC17226;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 21301.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

PF 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

```
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 21301; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX cDNA expression and secretion vectors.
XX Sequence 444 BP; 145 A; 88 C; 91 G; 120 T; 0 other;
```

alignment_scores:

Quality: 640.50 Length: 138

Ratio: 4.744 Gaps: 3

Percent Similarity: 97.826 Percent Identity: 97.101

alignment_block:

US-09-685-696-176 x AAC17226 ..

Align seg 1/1 to: AAC17226 from: 1 to: 444

228 LysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIl 244

:::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

38 AGAATCGATGTCACCCGTAAGAAATTCGCGGCTCTCGAGAAGTCGAT 87

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

244 eThrIleLeuSerThrProGluGlyThrSerAlaAlaCysLysSerIleL 261

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

88 TACTATCCTCTCTACTCTCTGAAGGCACCTCTGCGGCTTGTAACTCTATTC 137

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

261 euGluIleMetHisLysGluAlaGlnAspIleLysPheThr.GluGluIl 277

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

138 TGGAGATTATGCATAGGAAGCTCAAGATATAAAATTCACAAGAG...AT 184

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

277 eProLeuLysIleIleuAlaHisAsnAsnPheValGlyArgLeuIleGlyL 294

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

185 CCCCTTGAAGATTTAGCTCATATAACTTTGTTGGACGCTCTATTGGTA 234

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

294 ySGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIle 310

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

235 AAGAAGGAAGAATCTTAAAAAATTTGAGCAAGACACAGACATAAATC 284

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

311 ThrIleSerProLeuGlnGluLeuThrLeuTyRAsnProGluArgThrIl 327

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

285 ACGATATCTCCATTGCGAGGAATTTGACGCTGTATAATCCAGAACGCACAT 334

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

327 eThrValLysGlyAsnValGluThrCys.AlalysAlaGluGluIle 343

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

335 TACAGTTAAAGGCAATGTTGAGACATGTTGCCAAAGCTGAGGAGGATC 384

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

344 MetLysLysIleArgGluSerTyRAsnAspIleAlaSerMetAsnLe 360

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

385 ATGAAGAAATCAGGGAGTCTTTATGAAAATGATATTGCTTCTATGAATCT 434

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

360 uGlnAlaHis 363

|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

435 TCAAGCACAT 444

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI43398

seq_documentation_block:

ID AAI43398 standard; DNA; 588 BP.

XX AC AAI43398;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #12084 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX OS genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID No 12084; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 588 BP; 156 A; 139 C; 143 G; 150 T; 0 other;

alignment_scores:

Quality: 480.00 Length: 95

Ratio: 5.053 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:

US-09-685-696-176 x AAI43398 ..

Align seg 1/1 to: AAI43398 from: 1 to: 588

134 GlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPhetrLeuLy 150

19 AGAGCACTAGACAAACCTGAATGGATTTCAGTTAGAGAATTCACCTTGAA 68

150 sValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnG 167

69 AGTAGCCTATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGACG 118

167 lnProArgGlyArgGlyLeuGlyGlnArgGlySerArgGlyGlnGly 183

119 AGCCCCAGGTGCGCGGGGCTTGGGAGAGGGGCTCTCAAGGCAGGGG 168

184 SerProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLe 200

169 TCTCCAGGATCCGTATCCAAGCAGAAACCATGTGATTTGCCTCTGCGCT 218

200 uLeuValProThrGlnPheValGlyAlaIleGlyLysGlyGlyAlaI 217

219 GCTGGTTCCCAACCAATTTGTTGGAGCCATCATAGAAAAGAGTGCCA 268

217 hrIleArgAsnIleThrLysGlnThrGlnSerLys 228

269 CCATTCGAACATCCCAACAGACCCAGTCTAAG 303

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI56378

seq_documentation_block:

ID AAI56378 standard; DNA; 282 BP.

XX AC AAI56378;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #25064 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX OS genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID No 25064; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 282 BP; 77 A; 79 C; 70 G; 56 T; 0 other;

alignment_scores:

Quality: 474.00 Length: 93

Ratio: 5.097 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-685-696-176 x AAI56378 ..

Align seg 1/1 to: AAI56378 from: 1 to: 282

135 AlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysVa 151
|||||
2 GCACTAGACAAACTGAATGGATTTCAGTTAGAGAAATTCACCTTGAAGT 51
|||||
151 lAlaTyrlleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnInp 168
|||||
52 AGCCATATCCCTGATGAATGGCCGCCAGCAAAACCCCTTGCAGCAGC 101
|||||
168 rAArgGlyArgGlyLeuGlyGlnArgGlySerSerArgGlnGlySer 184
|||||
102 CCCGAGGTCCCGGGGCTTGGCAGAGGGGCTCTCAAGGCAGGGGTCT 151
|||||
185 ProGlySerValSerLysGlnLysProCysAspLeuProLeuArgLeuLe 201
|||||
152 CCAGGATCCGTATCCCAAGCAGAAACCATGTGATTTGCCCTCTCGCCCTGCT 201
|||||
201 uValProThrGlnPheValGlyAlaIleIleGlyLysGlnGlyAlaThrI 218
|||||
202 GGTTCACCCCAATTTGTTGGAGCCATCATAGAAAAGAAAGGTGCCACCA 251
|||||
218 leArgAsnIleThrLysGlnThrGlnSer 227
|||||
252 TTCGGAACATCACCAAAACAGACCCAGTCT 280
|||||

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC03267

seq_documentation_block:
ID AAC03267 standard; cDNA: 364 BP.

AC AAC03267;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3265.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GBST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG03261.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 3265; 7lpp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 364 BP; 84 A; 101 C; 118 G; 59 T; 2 other;

alignment_scores:
Quality: 341.00 Length: 96
Ratio: 4.060 Gaps: 0
Percent Similarity: 87.500 Percent Identity: 64.583

alignment_block:

US-09-685-696-176 x AAC03267 ..

Align seg 1/1 to: AAC03267 from: 1 to: 364

1 MetAsnLysLeuTyrlleGlyAsnLeuSerGluAsnAlaAlaProSerAs 17
|||||
76 ATGAACAAGCTTTACATCGGGAACCTGAGCCCGCGCTCACCGCCGACGA 125
|||||

17 pLeuGluSerIlePheLysAspAlaLysIleProValSerGlyProPheL 34
|||||
126 CCTCCGCAGCTCTTTGGGACAGGAAGCTGCCCTGGCGGACAGGTCC 175
|||||

34 euValLysThrGlyTyrlleAlaPheValAspCysProAspGluSerTrpAla 50
|||||
176 TGCTGAAGTCCGGCTAGCGCTTCGTGGACTACCCGACCAAGAACTGGGCC 225
|||||

51 LeuLysAlaIleGluAlaLeuSerGlyLysIleGluLeuHisGlyLysPr 67
|||||
226 ATCCGCGCATCGAGACCCCTCTCGGGTAAAGTGAATTCATGGGAAAT 275
|||||

67 oileGluValGluHisSerValProLysArgGlnArgIleArgLysLeuG 84
|||||
276 CATGGAGTGTATTACTAGTCTCTAAAAGCTAAGCAGCAGGAAATTC 325
|||||

84 InIleArgAsnIleProProHisLeuGlnTrpGluVal 96
|||||
326 AGATTGGAACATCCCTCCTCACCTGCAGTGGGAGGTG 363
|||||

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT: AAH31011

seq_documentation_block:

ID AAH31011 standard; cDNA: 424 BP.

AC AAH31011;

DT 27-JUL-2001 (first entry)

DE Human colon cancer cell line Kml2L4-A cDNA library derived sequence #945.

KW Human; diagnosis: colon cancer; cancer; malignant; chromosome mapping;
KW detection; colon cancer cell line Kml2L4-A; ss.

OS Homo sapiens.

PN WO200018916-A2.

PD 06-APR-2000.

PF 23-SEP-1999; 99WO-US22226.

PR 28-SEP-1998; 98US-0102161.

PR 28-SEP-1998; 98US-0102180.

PR 29-SEP-1998; 98US-0102380.

PR 08-OCT-1998; 98US-0103815.

PR 27-OCT-1998; 98US-0105877.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 DR WPI; 2000-293155/25.
 XX
 XX Polynucleotide library comprising 1079 defined sequences, useful in
 PT the form of an array to detect cancer or susceptibility to cancer -
 PT
 XX Claim 1; Page 454; 502pp; English.
 XX The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
 CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
 CC specifically binds to (II); (5) a vector comprising (1); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences
 CC are useful for chromosome mapping and detection of transcription levels.
 CC The 1079 polynucleotide sequences were derived from a human colon cancer
 CC cell line Kml21L4-A cDNA library.
 XX
 XX Sequence 424 BP; 117 A; 106 C; 111 G; 90 T; 0 other;

alignment_scores:
 Quality: 319.50 Length: 104
 Ratio: 3.590 Gaps: 1
 Percent Similarity: 85.577 Percent Identity: 57.692

alignment_block:

US-09-685-696-176 x AAH31011

Align seg 1/1 to: AAH31011 from: 1 to: 424

```

81 ArgLysLeuGlnIleArgAsnIleProHisLeuGlnTrpGluValLe 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 AGGAAATTCAGATCGAATCGAACATCCCTCCATCCTGCAGTGGGAGGT 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 uAspSerLeuLeuValGlnTyrGlyValValGluSerCysGluGlnValA 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 GGATGGACTTTGGCTCAATATGGAGAGTGGAGAAATGTGGAACAAGTCA 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 snThrAspSerGluThrAlaValValAsnValThrTyrSerSerLysAsp 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 ACACAGACACAGAAACCGCGCTTGTCAACGTCACATATGCAACAAGAGAA 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 GlnAlaArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPh 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 GAACGAAAAATTCAGATCGGAGAACTAAGCGGCATCATGTTTGAGAACTA 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 eThrLeuLysValAlaTyrIleProAspGluMetAlaAlaGlnGlnAsn 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 CTCCTTCAAGATTCCTACATCCCGGATGAAGAGGTGAGCTCCCTTCGC 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 RoLeuGlnGlnProArgGlyArgGlyLeuGlyGlnArgGlySerSer 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359 CCCCTCAGCGA.....GCCCAGCTGGGGACCAC 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 ArgGlnGlySer 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 TCTTCGGGAGC 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.AAA01526

seq_documentation_block:

ID AAA01526 standard; cDNA; 300 BP.

XX AAA01526;
 XX
 XX 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1517.
 XX
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09958675-A2.
 XX
 PD 18-NOV-1999.
 XX
 XX 13-MAY-1999; 99WO-0510602.
 PF
 XX 14-MAY-1998; 98US-0085426.
 PR
 PR 15-MAY-1998; 98US-0085537.
 PR
 PR 15-MAY-1998; 98US-0085596.
 PR
 PR 21-OCT-1998; 98US-0105234.
 PR
 PR 27-OCT-1998; 98US-0105877.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 PA (HYSE-) HYSEQ INC.
 XX
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX
 XX WPI; 2000-126369/11.
 DR
 XX Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PT
 XX Claim 1; Page 545; 1097pp; English.
 PS
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SQ Sequence 300 BP; 69 A; 95 C; 82 G; 54 T; 0 other;

alignment_scores:

Quality: 284.50 Length: 98
 Ratio: 3.428 Gaps: 3
 Percent Similarity: 84.694 Percent Identity: 60.204

alignment_block:

US-09-685-696-176 x AAA01526

Align seg 1/1 to: AAA01526 from: 1 to: 300

155 ProAspGluMetAlaAlaGlnGlnAsnProLeuGlnProArgGlyAr 171


```
||||||||| : : : : : : : : : : : ||| ||| : : : : :
17 CCGGATGAAGAGGTGAGCTCCCTTCGCCCTCAG.....CGAGCCCA 60
||| gArgGlyLeuGlyGlnArgGlySerArgGlnGly...SerProGlyS 187
: ||| : : : : : : : ||| : : ||||| : : ||||| :
61 GCGT.....GGGACCATCTCTCCGGGAGCAAGGCCAGCCCTTGGGG 104
187 erValSerLysGlnLysProCysAspLeuProLeuArgLeuLeuValPro 203
: : : : : : : : : ||| : : ||||| : : ||||| :
105 GCATTCTCAGGCCAGACAGATTGATTCCGCTCGGATCTCTGTCGCC 154
204 ThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAs 220
||||||||| : : : : : : : ||||| ||||| : : : : :
155 ACCCAGTTTGTGGTCCCATCATCGAAAGGAGGGCTTGACCATAAAGAA 204
220 nIleThrLysGlnThrGlnSerLysIleAspValHisArgLysGluAsnA 237
||||||||| : : : : : : : : : : : ||||| ||||| :
205 CATCACTAAGCAGACCCAGTCCCGGGTAGATATCCATAGAAAAGAACT 254
237 laGlyAlaAlaGluLysProCysIleThrIleLeuSerThrProGlu 251
: : ||||| : : : : : : : : : : : : : : :
255 CTGGAGCTGCAGAGAAGCCTGTCAACCATCCATGCCACCCACAGAG 298
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 14:55:39 ; Search time 65.7 Seconds
(without alignments)
652.792 Million cell updates/sec

Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAPSDLES.....VKHQHQKALQSGPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_1101.*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2956	100.0	579	21 AAB11328	Human lung cancer-
2	2943	99.6	579	21 AAB11365	Human lung cancer
3	2190	74.1	577	20 AAY30649	A murine c-myc cod
4	474	16.0	93	22 AAM38501	Peptide #12538 enc
5	341	11.5	97	21 AAG03261	Human secreted pro
6	241	8.2	644	15 AAR58813	Human c-myc far up
7	231.5	7.8	313	21 AAG08900	Arabidopsis thalia
8	229.5	7.8	590	15 AAR58816	Human c-myc far up
9	228.5	7.7	643	15 AAR58814	Human c-myc far up
10	227	7.7	45	22 AAM37315	Peptide #11352 enc
11	227	7.7	530	13 AAR28888	Ri paraneoplastic

12	217.5	7.4	644	21 AAG39346	Arabidopsis thalia
13	213.5	7.2	403	21 AAB42547	Human OREX ORF2311
14	213.5	7.2	403	22 AAB94825	Human protein sequ
15	204.5	6.9	781	21 AAG31213	Arabidopsis thalia
16	204.5	6.9	803	21 AAG31212	Arabidopsis thalia
17	204.5	6.9	846	21 AAG31211	Arabidopsis thalia
18	201.5	6.8	590	21 AAG39347	Arabidopsis thalia
19	201.5	6.8	1195	11 AAR05530	High density lipop
20	201.5	6.8	1292	11 AAR05531	High density lipop
21	191.5	6.5	419	21 AAG21062	Arabidopsis thalia
22	191.5	6.5	452	21 AAG21061	Arabidopsis thalia
23	190.5	6.4	462	22 AAB49335	Human hnRNPk prote
24	190	6.4	463	14 AAR37866	Rat liver catalase
25	190	6.4	463	20 AAY39467	Heterogenous ribon
26	190	6.4	464	14 AAR38906	Rat liver catalase
27	187	6.3	465	21 AAB43753	Human cancer assoc
28	187	6.3	465	22 AAG73800	Human colon cancer
29	181.5	6.1	446	21 AAG39348	Arabidopsis thalia
30	172	5.8	495	21 AAG39951	Arabidopsis thalia
31	172	5.8	500	21 AAG39950	Arabidopsis thalia
32	170.5	5.8	629	21 AAG36715	Arabidopsis thalia
33	167.5	5.7	228	21 AAG08901	Arabidopsis thalia
34	159.5	5.4	229	21 AAG21063	Arabidopsis thalia
35	159.5	5.4	471	22 AAB30930	Amino acid sequenc
36	156.5	5.3	571	21 AAG36716	Arabidopsis thalia
37	154.5	5.2	639	21 AAG40000	Arabidopsis thalia
38	154	5.2	343	13 AAR28887	pR18 fusion protei
39	153.5	5.2	230	17 AAR95954	Eukaryotic cell gr
40	151.5	5.1	418	21 AAG39952	Arabidopsis thalia
41	150	5.1	623	19 AAW68095	Chlamydomonas rein
42	147	5.0	243	15 AAR58815	Human c-myc far up
43	146	4.9	650	21 AAY85180	Cellulose synthase
44	145.5	4.9	414	17 AAW10529	Saccharomyces cere
45	145.5	4.9	414	22 AAB30809	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB11328
ID AAB11328 standard; Protein; 579 AA.
XX
AC AAB11328;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated protein L523S.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
(CORI-) CORIXA CORP.
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
DR N-PSDB; AAC65900.
XX

PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -

PS Claim 3; Page 186-188; 261pp; English.

XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2 and then administered to the patient to inhibit
XX development of cancer.

SQ Sequence 579 AA;

Query Match 100.0%; Score 2956; DB 21; Length 579;
Best Local Similarity 100.0%; Pred. No. 4.9e-246;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKLYIGNLSENAPDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALSGK 60
DB 1 mnklyignlsenapsdlesifkdakipvsgpfivktgyafvdcpdsewalkatealsgk 60
QY 61 IELHGKPIEVHSEVPRQRIRKQIRNIPPHLQWEVLDSLVOYGVVESCSEQVNTDSETA 120
DB 61 ielhgkpievehsvprqrirkqirnipphlqwevlidslvgyvvesceqvntdseta 120
QY 121 VNVVYSSKQDQARQALDKLNGFQLENFTLKVAYIPDEMAAQNPLOQPRRGLGQRGSS 180
DB 121 vnvvtsskqdarqaldklngfqlenftlkvayipdemaagnploqprrrglqrgss 180
QY 181 RQSPGSVSKQKPCDDLPLRLVPTQFVGALIGEGATIRNITKQTSKIDVHRKENAGAA 240
DB 181 rgspgsvsqkpcddlplrlvptqfvgaligegatirnitkqtskldvhrkenagaa 240
QY 241 EKSITILSTPEGTSAAKCSILEIMHKEAQDIKFTTEEIPKLAHNNFVGRGLGKEGRNLK 300
DB 241 eksitilstpegtsaacksileimhkeaqdikfteeteeiplkiahnnfvgrlqgkgrnlk 300
QY 301 KIEQDQDTKITISPLQELTLYNPERTITVKGNETCAKABEEIMKKIRESYENDIASMNL 360
DB 301 kieqdtktitispqlqeltlynpertitvkgnetcakaeeimkkiresyendiasmnl 360
QY 361 QAHLLIFGLNUNALGLPPTSGMPPPTSGPPSAMTPPYQPEQSETETVHQFIPALSVGAI 420
DB 361 qahllipglunlalnaglfpptsgmppptsgppsamtpypyqfegsetetvqhfpalsvgai 420
QY 421 IGKOGQHIKOLSRFAGASIKIAEAPDAKVRVVIITGPPEAQFKAQGRYIGKIKENFV 480
DB 421 igkqgqhikolsrfagasi kieapadavrvmviltgppeaqfkaqgriygkikeenf 480
QY 481 SPKEEVKLEAHIRVPFAAGRVIGKGGKTVNELQNLSAEVVPVPRDQTPDENQDVVVKIT 540
DB 481 spkeevkleahirvpfaagrvi gkggtvnelqnlsaevvprvprdqtpdenqdvvvkit 540
QY 541 GHFYACQVARKTQEILTQVKKHQKQKALQSGPPQSRKK 579
DB 541 ghfyacvarktqeiltqvkkhqkqkalqsgppqsrkk 579

RESULT 2

AAB11365

ID AAB11365 standard; Protein; 579 AA.

XX

XX AAB11365;

XX

DT 21-FEB-2001 (first entry)

XX

XX Human lung cancer associated antigen L523S.

XX

KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection.

OS Homo sapiens.

XX WO200061612-A2.

XX 19-OCT-2000.

XX 03-APR-2000; 2000WO-US08896.

XX 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX (CORI-) CORIXA CORP.

XX Wang T, Fan L;

XX WPI; 2000-628399/60.

DR N-PSDB; AAC66035.

XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX protein is used for detecting and monitoring progression of lung cancer
XX in a patient -

XX Claim 3; Page 259-261; 261pp; English.

XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2 and then administered to the patient to inhibit
XX development of cancer.

XX Sequence 579 AA;

Query Match 99.8%; Score 2943; DB 21; Length 579;
Best Local Similarity 99.7%; Pred. No. 6.5e-245;
Matches 577; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNKLYIGNLSENAPDLESIFKDAKIPVSGPFLVKTGYAFVDCPDSEWALKATEALSGK 60

DB 1 mnklyignlsenapsdlesifkdakipvsgpfivktgyafvdcpdsewalkatealsgk 60

QY 61 IELHGKPIEVHSEVPRQRIRKQIRNIPPHLQWEVLDSLVOYGVVESCSEQVNTDSETA 120

DB 61 ielhgkpievehsvprqrirkqirnipphlqwevlidslvgyvvesceqvntdseta 120

QY 121 VNVVYSSKQDQARQALDKLNGFQLENFTLKVAYIPDEMAAQNPLOQPRRGLGQRGSS 180

DB 121 vnvvtsskqdarqaldklngfqlenftlkvayipdemaagnploqprrrglqrgss 180

QY 181 RQSPGSVSKQKPCDDLPLRLVPTQFVGALIGEGATIRNITKQTSKIDVHRKENAGAA 240

DB 181 rgspgsvsqkpcddlplrlvptqfvgaligegatirnitkqtskldvhrkenagaa 240

QY 241 EKSITILSTPEGTSAAKCSILEIMHKEAQDIKFTTEEIPKLAHNNFVGRGLGKEGRNLK 300

DB 241 eksitilstpegtsaacksileimhkeaqdikfteeteeiplkiahnnfvgrlqgkgrnlk 300

QY 301 KIEQDQDTKITISPLQELTLYNPERTITVKGNETCAKABEEIMKKIRESYENDIASMNL 360

```
Db 301 klegtdtkitispilqetllynpertitvkgvnetcakaeeimkkiresyendiasmnl 360
Qy 361 QAHLIPGLNALGFLPPTSGMPPTSGPPSAMTPYPQFQSETEVHQRIPALSVGAI 420
Db 361 qahlipglnlnalglfpptsgmpptsgppsamcpypqfseqsetcvnlhlfipalsvgai 420
Qy 421 IGKQGHKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPAEQAKGRIYKIKKEENFV 480
Db 421 lggqgghikqlsrfagasiklapeapdakvrmvliitgppaeqfkaqgriykgkikeenf 480
Qy 481 SPKEEVKLEAHIRVPSFAAGRVIGKGTVNELQNLSSAEVVPDRDQPDENDQVVKIT 540
Db 481 spkeevkleahirvpsfaagrvigkgtvnelqnlssaevvprdrqtpdendqvvkkit 540
Qy 541 GHFYACQVAQRKIQEILTVQKHQOQKALQSGPPQSRRK 579
Db 541 ghfyacvraqrkiqeilcvkhqgqkqalqsgppqsrirk 579

RESULT 3
ID AAY30649 standard; Protein; 577 AA.
XX
AC AAY30649;
XX
DT 17-NOV-1999 (first entry)
XX
DE A murine c-myc coding region determinant binding protein.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor;
KW c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer.
XX
OS Mus musculus.
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US04897.
XX
PR 09-MAR-1998; 98US-0077372.
XX
PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
XX
DR N-PSDB; AAZ10617.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
XX coding region determinant-binding protein
XX
FS Example; Fig 1A-D; 79pp; English.
XX
CC The present sequence represents a murine c-myc coding region determinant
XX binding protein (CRD-BP). The presence or absence of a tumor can be
XX determined by determining the levels of CRD-BP present in the suspect
XX tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
XX and so prolongs its half-life. The methods are used for diagnosing
XX presence or absence of a tumor in a human, especially breast, colon
XX and pancreatic cancer. They are also used to inhibit cancer cell
XX growth.
XX
SQ Sequence 577 AA;

Query Match 74.1%; Score 2190; DB 20; Length 577;
Best Local Similarity 74.1%; Pred No. 5.2e-180;
Matches 434; Conservative 62; Mismatches 74; Indels 16; Gaps 8;

Qy 1 MNKLYIGNLSNAAPSDLESIFDKAIPVSGFFLVKTYGAFVDCPDSEWALKATEALSGK 60
```

```
Db 1 mnklyignlsnavtpadlekvaehkisyggflvksyafvdcpdchamkaiefsgk 60
Qy 61 IELHGKPIEVEHSVKKRQIRKLOIRNIPPHLOWEVLDSLLVQYGVVSECEQVNTSETA 120
Db 61 velqgkrlemehsvpkkgrsrkigrnippqirwvldslaqygtvenceqvnteseta 120
Qy 121 VYNVYSSKQDARQALDKLNGFQLENFTLKVAYIPDEMAAQONPLQOPRRR-GLQGRGS 179
Db 121 vnyvysnregrtqaimkinghqlenhalksyipdeqitq---gpengrrggfgrsq 176
Qy 180 SRQSGP---GSVSKQKPCDPLRLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKEN 236
Db 177 prqgspvaagapakgqvdiplrlivptqyvgailgkegatiirnicktqskidvhrken 236
Qy 237 AGAAEKSTILTPBETSAACKSILIEIMHKEAQDIKFTTEIPLKILAHNNFVRLIGKEG 296
Db 237 agaaekaisvhtpegcsackmleimhkeakdtktadevpkilaahnfvrligk 296
Qy 297 RNLKKIEQDTRKITISPLQELTLNPERTITVKGNETCAKAEIEIMKKIRESYENDIA 356
Db 297 rnlkveqdtetkitisslqdlitlynpertitvkgaiencraeqeimmkkyreayenda 356
Qy 357 SMNLAHLIPGLNALGFLPPTSGMPPTSGPPSAMT--PPYQFEQS-ETETVHQFIP 413
Db 357 amslqshlipglnlaavgilpassavvp---ppssvtgaapyssfmqapeqemvqvfi 413
Qy 414 ALSVCAIIGKQGHKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPAEQAKGRIYK 473
Db 414 aqvagailgkqghikqlsrfagasiklappetpdskvrmvviitgppaeqfkaqgriy 473
Qy 474 IKEENFVSPKEVKLEAHIRVPSFAAGRVIGKGTVNELQNLSSAEVVPDRDQPDEND 533
Db 474 lkeenffpkeevkleahirvpsaagrvigkgtvnelqnlssaevvprdrqtpdend 533
Qy 534 QVVVKITGHFYACQVAQRKIQEILTVQKHQOQKALQSGPPQSRRK 579
Db 534 qvivkiighfyasqmagrkirdilqkvkq-qhqqkq-qsnaqarrk 577

RESULT 4
AAM38501
ID AAM38501 standard; Protein; 93 AA.
XX
AC AAM38501;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #12538 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, .Hauzel DK, Chen W, Rank DR;
XX
PI Penn SG, .Hauzel DK, Chen W, Rank DR;
XX
```

DR WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
PS Claim 27; SEQ ID No 38770; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 93 AA;
SQ
Query Match 16.0%; Score 474; DB 22; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.6e-33;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 135 ALQKNGFQLENFTLVKVAIPDEMAAQNPLQPRRRGLGQSGSRQSGSVSKOKPC 194
Db 1 aldkngfqlenftlvkvaipdemaaqnplqprrrglgqsgsrqsgsvskokpc 60
QY 195 DLPLRLLVPTQFVGAIGKEGATIRNTKQTQS 227
Db 61 dlplrlvlptqfvgaiigkegatinrntkqtqs 93
RESULT 5
AAG03261
ID AAG03261 standard; Protein; 97 AA.
XX
AC AAG03261;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 7342.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR N-PSDB; AAC03267.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 7342; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5',
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 97 AA;
SQ
Query Match 11.5%; Score 341; DB 21; Length 97;
Best Local Similarity 64.6%; Pred. No. 7.9e-22;
Matches 62; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
QY 1 MNKLYIGNLSENAAPSDELSIFKDAKIPVSGPFLVKTGYAFVDCPDDESWALKATEALSGK 60
Db 2 mnklyignlspavtaddrlqfgrdkipiaqgvlksgafvdyppdqnwairatetisgk 61
QY 61 IELHGKPIEVESHVYPRKQIRKQLQIRNIPPHLOWEV 96
Db 62 velhgkimevdysvskklsrskiqirnipphlqwev 97
RESULT 6
AAR58813
ID AAR58813 standard; Protein; 644 AA.
XX
AC AAR58813;
XX
DT 13-APR-1995 (first entry)
XX
DE Human c-myc far upstream element (FUSE) binding protein (FBP).
XX
KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW promoter Pl.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 149 /label= M,I
XX
PN W09419465-A.
XX
PD 01-SEP-1994.
XX
PF 22-FEB-1994; 94WO-US01782.
XX
PR 22-FEB-1993; 93US-0021608.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Avigan MI, Duncan RC, Levens DL;
XX
DR WPI; 1994-294330/36.
DR N-PSDB; AAQ68909.
XX
PT New DNA-binding regulator of c-myc expression and its cDNA - used
PT to develop prods. for diagnosis and therapy of disease states
PT such as tumour formation
XX
XX Claim 5; Page 40-43; 94pp; English.
XX
CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
CC upstream element) binding protein (FBP) are synonymous. A FUSE which
CC is required for maximal transcription of c-myc binds a factor (DROME
CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC human c-myc promoter Pl. A full length FBP cDNA sequence was
CC assembled from overlapping clones obtd. from cDNA libraries. Source
CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC BJAB, and PMAP/PHA stimulated pooled human peripheral blood
CC lymphocytes (PBLS). A composite cDNA is given in AAQ68909 and the
CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in AAR58813.

XX	Sequence	644 AA;
XX	Query Match	8.2%; Score 241; DB 15; Length 644;
XX	Best Local Similarity	22.3%; Pred. No. 6.1e-12;
XX	Matches 100; Conservative	75; Mismatches 154; Indels 120; Gaps 17;
QY	161 QQNPLQ---QPRGRRGLGQRSGSSQKQPCDLPRLRLVPTQFVGAIIGKEGAT	217
DB	161 QQNPLQ---QPRGRRGLGQRSGSSQKQPCDLPRLRLVPTQFVGAIIGKEGAT	217
DB	63 qkpledgdpdakkvapqndsfqtqlppmhqgqsrvmteeykvpdgmvgfiigrageq	122
QY	218 IRNITKQTSKIDVHRKENAGAAEKSTIILSTPBGTSAAKCSILEIMHKEAQDITKFT--	275
DB	218 IRNITKQTSKIDVHRKENAGAAEKSTIILSTPBGTSAAKCSILEIMHKEAQDITKFT--	275
DB	123 isriqgesgcklq1-apdsdglpsercxilgtpesvgsakrllldqivekgtrpapgfhgd	181
QY	276 --EIPKILAHNNFVRLGCKEGRNLKKTEODTDTTITTSPLQELTYLNPERT-----I	327
DB	276 --EIPKILAHNNFVRLGCKEGRNLKKTEODTDTTITTSPLQELTYLNPERT-----I	327
DB	182 gpnavaqeimipaskagdlvlgkggetikqlqeragvkmvm--lqd---gpnctgadkpl	235
QY	328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASMNLQAHILPGLNLALGLPPTSGM	382
DB	328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASMNLQAHILPGLNLALGLPPTSGM	382
DB	236 ritgdpykvqakemvliirldqgdfrevrneygsr-----lqg-----negi	278
QY	383 PPTSGPPSAMPYPPOFEQSETEVTHQFTALSVGAIIGKQGHIIKOLSRFAGASTKIA	442
DB	383 PPTSGPPSAMPYPPOFEQSETEVTHQFTALSVGAIIGKQGHIIKOLSRFAGASTKIA	442
DB	279 dvp-----lprfavgivigrngemikkikqndagvriqfk	312
QY	443 PAEAPDAKVRVLIITGPE-AQFKAQ-----GRIYCK-----	473
DB	443 PAEAPDAKVRVLIITGPE-AQFKAQ-----GRIYCK-----	473
DB	313 pddgttpe-rlaqltgpddrcqhaaelitdlrsvqagnpgpggrgrgrggnwnmg	371
QY	474 ----IKENFVSPKPEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVVPRDQTP	529
DB	474 ----IKENFVSPKPEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSAEVVVPRDQTP	529
DB	372 ppgglqefnfi-----vpgkgtglgkggetiksisqsgsarielqnrppp	418
QY	530 DENDQV-VVKITGH-----FYACQVAQRKI	553
DB	530 DENDQV-VVKITGH-----FYACQVAQRKI	553
DB	419 nadpnmkflftgtpqgldyarqlieeeki	447
RESULT	7	
AA08900	AA08900 standard; Protein; 313 AA.	
XX	AA08900;	
AC	AA08900;	
XX	AA08900;	
DT	17-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 6617.	
DE	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS	EP1033405-A2.	
PN	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	28-APR-1999; 99US-0130891.	
PR	28-APR-1999; 99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	30-APR-1999; 99US-0132407.	
PR	04-MAY-1999; 99US-0132484.	
PR	05-MAY-1999; 99US-0132485.	
PR	06-MAY-1999; 99US-0132486.	
PR	07-MAY-1999; 99US-0132487.	
PR	07-MAY-1999; 99US-0132863.	
PR	14-MAY-1999; 99US-0134218.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	25-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136392.	
PR	28-MAY-1999; 99US-0136782.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	
PR	07-JUN-1999; 99US-0137724.	
PR	08-JUN-1999; 99US-0138094.	
PR	10-JUN-1999; 99US-0138540.	
PR	10-JUN-1999; 99US-0138847.	
PR	14-JUN-1999; 99US-0139119.	
PR	16-JUN-1999; 99US-0139452.	
PR	16-JUN-1999; 99US-0139453.	
PR	17-JUN-1999; 99US-0139492.	
PR	18-JUN-1999; 99US-0139454.	
PR	18-JUN-1999; 99US-0139455.	
PR	18-JUN-1999; 99US-0139456.	
PR	18-JUN-1999; 99US-0139457.	
PR		

PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145218;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147204;
PR	04-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	09-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;
PR	11-AUG-1999;	99US-0148319;
PR	12-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	13-AUG-1999;	99US-0148684;
PR	16-AUG-1999;	99US-0149368;
PR	17-AUG-1999;	99US-0149375;
PR	18-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149722;
PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149829;
PR	23-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	26-AUG-1999;	99US-0150566;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151438;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	13-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154399;
PR	20-SEP-1999;	99US-0154779;
PR	23-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157865;
PR	07-OCT-1999;	99US-0158029;
PR	08-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;

PR	14-OCT-1999;	990S-0159330;
PR	14-OCT-1999;	990S-0159331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-0159584.
PR	21-OCT-1999;	990S-0160744.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.
PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	21-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160810.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.

Query Match 7.8%; Score 231.5; DB 21; Length 313;

Best Local Similarity	25.1%	Pred. NO. 1.3e-11;		
Matches	79;	Conservative	62;	Mismatches 131;
			Indels	43;
			Gaps	10;

Qy	183	GPSGSVKQKP-----CDLP--LRLLVPTQFGVAILIGEGATIRNITKQTSKID	230
		: : : : : : : : : : : : : : : :	
Db	10	gspeakarspephdsseadaekpthrflvnaaagsvlgkgsttlfqaksgariq	69
Qy	231	VHREN--AGAAEKSIITLSTPECTSAACKSILEIMHKEAQDIKFTEEP---LKILAHN	285
		: : : : : : : : : : : : : : : :	
Db	70	lsrqefpfgtdrlmimgskivevnglellldklhsehaedgneveprriirivpn	129
Qy	286	NFVGRLLGKGRNLKKTQODTKTITISPLQELTLN--PERTITVKGNVETCAKEEETM	344
		: : : : : : : : : : : : : : : :	
Db	130	sscgilgkggatkfsfieskagikispldn--tfyglrdrlvtlsgtfeeqmraidlll	188
Qy	345	KKIRESYENDIASMNLQAHILPGLNIALGLFPPTSGMPPTSGPPSATMPYPQEQSE	404
		: : : : : : : : : : : : : : : :	
Db	189	aklt---eddhysqnvhspysyaagynsvnyapngsg-----gkyqnkheea	232
Qy	405	TETVHQFIPALSYGAILGKOGOHIKOLSRFAGASIKIAP-----AEAPAKVVMWIICP	459
		: : : : : : : : : : : : : : : :	
Db	233	stvtvigaadehglvlvgrgrnrmelqmtgarikisdrgfmsgttdrkvs---itgp	289
Qy	460	PEAQFKAQGRYICKI	474
		: : : : : : : : : : : : : : : :	
Db	290	qraiqgaetmikav	304

RESULT	8
AAR58816	
ID	AAR58816 standard; Protein; 590 AA.
XX	
XX	
AC	AAR58816;
XX	
XX	
DT	13-APR-1995 (first entry)
XX	
DE	Human c-myc far upstream element (FUSE) binding protein (FBP)
DE	variant from PBL clone 31-10.
DE	

XX PF 22-FEB-1994; 94WO-US01782.
 XX PR 22-FEB-1993; 93US-0021608.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Avigan MI, Duncan RC, Levens DL;
 XX XX WPI; 1994-294330/36.
 XX DR N-PSDB; AAQ68912.
 XX XX
 XX PT New DNA-binding regulator of c-myc expression and its cDNA - used
 XX PT to develop prods. for diagnosis and therapy of disease states
 XX PT such as tumour formation
 XX XX
 XX PS Claim 20; Page 53-56; 94pp; English.
 XX XX
 CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
 CC upstream element) binding protein (FBP) are synonymous. A FUSE which
 CC is required for maximal transcription of c-myc binds a factor (DROME
 CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
 CC human c-myc promoter Pl. A full length FBP cDNA sequence was
 CC assembled from overlapping clones obtd.from cDNA libraries. Source
 CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
 CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
 CC lymphocytes (PBLs). A composite cDNA is given in AAQ68909 and the
 CC deduced polypeptide sequence. (67.5 kd mol. wt.) is given in AAQ58813.
 CC The clone from an activated PBL cDNA library labeled "31-10" contains
 CC an FBP variant. The ORF comprised of bps 135-1991 of AAQ68909 with
 CC two exceptions. Clone 31-10 contains 63 bps inserted at posn. 238
 CC which probably result from an intron which had not been spliced
 CC out. The inserted bps remain in the ORF. The 31-10 clone also
 CC deviates in that bps 1807-19522 are deleted. This deletion shifts
 CC out of frame the stop codons which would terminate translation in
 CC the other clones. The AA sequence for the protein encoded by clone
 CC 31-10 can be found in AAQ58816.
 XX XX
 XX SQ Sequence 590 AA;

Query Match 7.8%; Score 229.5; DB 15; Length 590;
 Best Local Similarity 22.5%; Pred. No. 5.2e-11;
 Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;

Qy 202 VPTQFVGAICGEGATIRNITKQTSKIDVHRKENAGAAEKISITILSPETGSACKSIL 261
 Db 91 vpdgmvgfllgrggedisrldqesgcklqi- apdsggiperscmllgtcpsvqakrllld 149
 Qy 262 RIMHKEAQDIKFTFTE-----EIPKTLAHNFFVRLIGKEGRNLKKEQDTDTKITISPLQ 316
 Db 150 qivekgrpapfhhgdpggnavqeimipaskagivlgkgtetikqlqragvkmvm--iq 207
 Qy 317 ELTYNPERT-----ITVKGNVETCAKAEETMKKIRE-----SYENDIASMNLQAHLP 366
 Db 208 d-----gpqntgadkplritgdykvqakenvlelirdqgfvreinygsr-----ig 257
 Qy 367 GLNLNALGLFPPTSCMPPTSGPSPATPPYPQPEQSETEVHVQFIPALSVCAIIGKGGQ 426
 Db 258 g-----negidvp-----iprfavgivigrnge 280
 Qy 427 HIKQLSRFAGASIKIAEAPADAKVRMVIITGPEE-AQKFAQ----- 467
 Db 281 mikkiqndagvriqkpdgttpe-riaqitgppdrcdhaeaitdlrrsvqagngpgpg 339
 Qy 468 --GRIYK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTVMNEL 513
 Db 340 pggfgrgrgqgnwmnmgppggigqefnfi-----vptgtglligkgetiksi 386
 Qy 514 QNLSSAEVVPRDQTPDENDOV-VVKITGH-----FYACQVAQRKI 553
 Db 387 sqqsgarieldrnpnpadpnmkiftirtgtpqgidyarqlleeki 431

RESULT 9
 AAR58814
 ID AAR58814 standard; Protein; 643 AA.
 XX AC AAR58814;
 XX DT 13-APR-1995 (first entry)
 XX XX Human c-myc far upstream element (FUSE) binding protein (FBP)
 XX DE variant lacking Ser 97.
 XX DE
 XX KW DNA-binding regulator; c-myc; DROME; FUSE; activator; cis-element;
 XX KW promoter Pl.
 XX XX Homo sapiens.
 XX OS
 FH Key Location/Qualifiers
 FT Misc-difference 148
 FT FT /label= M,I
 XX PN W09419465-A.
 XX PD 01-SEP-1994.
 XX PF 22-FEB-1994; 94WO-US01782.
 XX PR 22-FEB-1993; 93US-0021608.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Avigan MI, Duncan RC, Levens DL;
 XX XX WPI; 1994-294330/36.
 XX DR N-PSDB; AAQ68910.
 XX PT New DNA-binding regulator of c-myc expression and its cDNA - used
 XX PT to develop prods. for diagnosis and therapy of disease states
 XX PT such as tumour formation
 XX PS Claim 12; Page 60-63; 94pp; English.
 XX XX
 CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
 CC upstream element) binding protein (FBP) are synonymous. A FUSE which
 CC is required for maximal transcription of c-myc binds a factor (DROME
 CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
 CC human c-myc promoter Pl. A full length FBP cDNA sequence was
 CC assembled from overlapping clones obtd.from cDNA libraries. Source
 CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
 CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
 CC lymphocytes (PBLs). A composite cDNA is given in AAQ68909 and the
 CC deduced polypeptide sequence. (67.5 kd mol. wt.) is given in AAQ58813.
 CC Three clones from a BJAB cDNA library and three clones from an
 CC activated human PBL cDNA library contain the sequence in AAQ68909.
 CC Three clones from the same BJAB library and two clones from the
 CC activated lymphocyte library are lacking bps 316,317 and 318
 CC (see AAQ68910). The mRNA lacking these three bps would encode a
 CC variant protein lacking Ser 97 (see AAR58814).
 XX XX
 XX SQ Sequence 643 AA;

Query Match 7.7%; Score 228.5; DB 15; Length 643;
 Best Local Similarity 22.0%; Pred. No. 7.3e-11;
 Matches 99; Conservative 77; Mismatches 152; Indels 121; Gaps 18;

Qy 161 QONPLQ---QPRGRRLGQRSGSGSVSKQKPCDPLRLLLVPTQFVGAIGKEGAT 217
 Db 63 qkrpiedgdqpdakvapondsfgtqlp-pmhqqrsvmteeykvpdgmvgfllgrgqeq 121
 Qy 218 IRNITKQTSKIDVHRKENAGAAEKISITILSPETGSACKSILEIMHKEAQDIKFTFTE-- 275
 Db 122 lsriqgesgcklqi- apdsggiperscxltgtcpsvqskrllldgivekgrpagfhgd 180

XX	Sequence	530 AA:	XX	Sequence	530 AA:
XX	Query Match	7.7%; Score 227; DB 13; Length 530;	XX	01-APR-1999;	99US-0127452;
XX	Best Local Similarity	22.4%; Pred No. 7.3e-11;	XX	06-APR-1999;	99US-0128234;
XX	Matches 119; Conservative	77; Mismatches 179; Indels 156; Gaps 20;	XX	08-APR-1999;	99US-0128714;
QY	155 PDMAAQNPLQOPRGRRLGQGRSSQSGSVSKQKPCDPLRLVLPTQFVGAIICKE	214	XX	16-APR-1999;	99US-0129845;
DB	43 pd---srkrpleap-----peagstrktntgedgqyf----lkvlipsyaagslikgk	88	XX	19-APR-1999;	99US-0130077;
QY	215 GATYRNTTKQTSIDVHRKENA-----GAAEKSTILSPETSAAKCSILEIMHKEAQ	269	XX	21-APR-1999;	99US-0130449;
DB	89 gqtvlvqketgatikiskiskskdfpygttervcligtvealnnavhgfiaekirempq	148	XX	23-APR-1999;	99US-0130891;
QY	270 DIKTEEIPL-----KTLAHNNFVGR	290	XX	28-APR-1999;	99US-0131449;
DB	149 nvaktepvsilpqttvnpdrikqlpsptttksspsdpmttsranqvklivpnstagl	208	XX	30-APR-1999;	99US-0132048;
QY	291 LIKREGNLAKEODTQTKITISPLQELTYLNPRTITVKGNVETCAKAEIIMKKIRE-	349	XX	04-MAY-1999;	99US-0132485;
DB	209 iikggatvkavmeqsgaawqlsqkpd-ginlgervvtvsgepeqnkavelliqlikqed	267	XX	06-MAY-1999;	99US-0132486;
QY	350 -----SYEN---DIASMN-----LQAHLPGLNLNALGLFPPT-	379	XX	07-MAY-1999;	99US-0132863;
DB	268 pqsgscinisyantvgpansnptgsyantaevlpataaaaagllghanlagvaafpavl	327	XX	11-MAY-1999;	99US-0134256;
QY	380 SGMPPTSGPPSAMTPPYQPEQSETVTHQFIPALSVGAIIHGQGHKQLSRFAGASI	439	XX	14-MAY-1999;	99US-0134218;
DB	328 sgf---tgnldvait-----salntl-----asygynlntlglsqaaa-tgala	369	XX	14-MAY-1999;	99US-0134221;
QY	440 KIAPAEAPDAKVRVMIITGPPEAQFKAQGRYIGKI-----	474	XX	18-MAY-1999;	99US-0134768;
DB	370 aaaaanpaaaaaanlatyasea---sasgstagtagtalfglsaaataatngyfgaasp	427	XX	20-MAY-1999;	99US-0134941;
QY	475 -----KEENFVSPKEEVKLEAHIRVPFPAAGRVIGKGGKTVNELQNLSAEEVVP-	525	XX	21-MAY-1999;	99US-0135124;
DB	428 laasailgtekstgskdvve---iavpnlvgailgkgktiveyqellgariqisk	483	XX	25-MAY-1999;	99US-0135353;
QY	526 -DQTPDENQVVKVITGHFACVQAORKIQIELTQVQKHQQKALQSGPPQ	575	XX	27-MAY-1999;	99US-0136021;
DB	484 gefvgptrnrkvt-itgtpaatgaqyliitqrit-----yeggvraanpq	527	XX	28-MAY-1999;	99US-0136392;
RESULT	12		XX	01-JUN-1999;	99US-0136782;
AAG39346			XX	03-JUN-1999;	99US-0137222;
ID	AAG39346 standard; Protein; 644 AA.		XX	04-JUN-1999;	99US-0137502;
XX			XX	07-JUN-1999;	99US-0137724;
AC	AAG39346;		XX	08-JUN-1999;	99US-0138094;
DT	18-OCT-2000 (first entry)		XX	10-JUN-1999;	99US-0138540;
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 48669.		XX	10-JUN-1999;	99US-0138847;
XX	Protein identification; signal transduction pathway; metabolic pathway;		XX	14-JUN-1999;	99US-0139119;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		XX	16-JUN-1999;	99US-0139452;
KW	termination sequence.		XX	16-JUN-1999;	99US-0139453;
KW			XX	17-JUN-1999;	99US-0139492;
XX	Arabidopsis thaliana.		XX	18-JUN-1999;	99US-0139454;
OS	EPI033405-A2.		XX	18-JUN-1999;	99US-0139455;
XX	06-SEP-2000.		XX	18-JUN-1999;	99US-0139456;
XX	25-FEB-2000; 2000EP-0301439.		XX	18-JUN-1999;	99US-0139457;
XX	25-FEB-1999; 99US-0121825.		XX	18-JUN-1999;	99US-0139458;
PR	05-MAR-1999; 99US-0123180.		XX	18-JUN-1999;	99US-0139459;
PR	09-MAR-1999; 99US-0123548.		XX	18-JUN-1999;	99US-0139461;
PR	23-MAR-1999; 99US-0125788.		XX	18-JUN-1999;	99US-0139462;
PR	25-MAR-1999; 99US-0126264.		XX	18-JUN-1999;	99US-0139463;
PR	29-MAR-1999; 99US-0126785.		XX	18-JUN-1999;	99US-0139750;
XX			XX	18-JUN-1999;	99US-0139763;
XX			XX	21-JUN-1999;	99US-0139817;
XX			XX	22-JUN-1999;	99US-0139899;
XX			XX	23-JUN-1999;	99US-0140353;
XX			XX	23-JUN-1999;	99US-0140354;
XX			XX	24-JUN-1999;	99US-0140695;
XX			XX	24-JUN-1999;	99US-0140823;
OS			XX	29-JUN-1999;	99US-0140991;
XX			XX	30-JUN-1999;	99US-0141287;
PN			XX	01-JUL-1999;	99US-0141842;
XX			XX	01-JUL-1999;	99US-0142055;
XX			XX	02-JUL-1999;	99US-0142390;
PD					

PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0158293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 20-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149722.		
PR 20-AUG-1999;	99US-0149723.		
PR 23-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149902.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150560.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		
PR 24-SEP-1999;	99US-0155659.		
PR 28-SEP-1999;	99US-0156458.		
PR 29-SEP-1999;	99US-0156596.		

Query Match

Best Local Similarity

Matches

93;

Conservative

82;

Mismatches

177;

Indels

127;

Gaps

15;

QY 199	RLVPTQFVGAIGKEGATIRNITKOTSKIDVHRKENAGAAEKSTITLSTPEGTSAAK	258
Db 282	rlcpidvvgvgigsgkvinairhntkakikvfdqlh-gcsqrvtiycs-----	331
QY 259	SILEIMHKEAQDIKT--EEIPL-----	281
Db 332	-----vkeqeeigftksenepiccaqdallkvydaivasdeenctknvdrdnkcerl	386
QY 282	LAHNFVGRIGKEGRNLKIEQDUTKI-----TISPIQELTYLPERTITVKGNETC	336
Db 387	lvpfsgsssligkagenikrirttrasvkvsdkvdspsvhcamedynvvvisgepesv	446
QY 337	AK---AEEETMKKI--RESYENDIASMNL-----	364
Db 447	kqalfavsaimykinprenipidstsqdvpasviypsdlsnsvypqtgfyfynqdhllq	506
QY 365	---IPGLNINLGL-----FPPTSGMPPTSGPPSAMTPPYQFQESQETETVHQFIPALS	416
Db 507	gagvpsy-fnalsvsdfgyaetaanpvvfasslpythgfggssrseelvfkvlclpcn	565
QY 417	VGAIGKQGOHIKQLSRFACASIKIAPAPADAKVRWVIITGPPEAQF-----	464
Db 566	imrvigkggstikrيرهagscilevndsrctkcgddceviivtateillfcclstpfvfmg	625
QY 465	---KAQGRYKIKEENFVSPEEYKLEAHIRVPFPAAGRVIGKGTVNEQLNLSAE	520
Db 626	pddmksma-veavlllqeyindedaenvkmqlilvsskvgicvkgsgsvineirktnan	684
QY 521	VVVPDRDQTPDNDQVVVVKITGHFYACQVAKRQKIQEILLTOVKHQKQKALQSGPPQSRK	579
Db 685	iciskgkdd-----lvevsgev-----ssvrdaliqivlrredvlgdkdsvatr	731

Search completed: January 25, 2002, 22:11:55
Job time: 26176 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 15:00:14 ; Search time 247.67 Seconds
(without alignments)
52.608 Million cell updates/sec

Title: US-09-685-696-176

Perfect score: 2956

Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKHQHQKALQSGPPQSRRK 579

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pap.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2190	74.1	577	4	US-09-261-855-2
2	241	8.2	644	1	US-08-021-608D-2
3	241	8.2	644	1	US-08-726-160-2
4	241	8.2	644	5	PCT-US94-01782-2
5	238	8.1	49	4	US-09-261-855-22
6	237	8.0	48	4	US-09-261-855-24
7	236.5	8.0	643	1	US-08-021-608D-10
8	236.5	8.0	643	1	US-08-726-160-10
9	236.5	8.0	643	5	PCT-US94-01782-10
10	232	7.8	49	4	US-09-261-855-18
11	229.5	7.8	590	1	US-08-021-608D-8
12	229.5	7.8	590	1	US-08-726-160-8
13	229.5	7.8	590	5	PCT-US94-01782-8
14	227	7.7	47	4	US-09-261-855-23
15	227	7.7	530	1	US-08-187-793-4
16	225	7.6	47	4	US-09-261-855-21
17	219	7.4	48	4	US-09-261-855-20
18	218	7.4	47	4	US-09-261-855-17
19	190	6.4	47	4	US-09-261-855-19
20	159.5	5.4	471	4	US-08-866-928B-1
21	154	5.2	343	1	US-08-187-793-2
22	147	5.0	243	1	US-08-021-608D-6
23	147	5.0	243	1	US-08-726-160-6
24	147	5.0	243	5	PCT-US94-01782-6
25	146.5	5.0	688	4	US-08-973-273-26
26	146.5	5.0	747	4	US-08-973-273-3
27	145.5	4.9	414	1	US-07-667-276A-4

28	140.5	4.8	655	4	US-09-347-833-4	Sequence 4, Appli
29	139	4.7	444	1	US-07-881-075-3	Sequence 3, Appli
30	139	4.7	444	1	US-08-120-827-3	Sequence 3, Appli
31	139	4.7	444	1	US-08-478-675-3	Sequence 2, Appli
32	133.5	4.5	359	1	US-07-881-075-2	Sequence 2, Appli
33	133.5	4.5	359	1	US-08-120-827-2	Sequence 2, Appli
34	133.5	4.5	359	1	US-08-478-675-2	Sequence 2, Appli
35	133	4.5	380	1	US-07-881-075-51	Sequence 51, Appl
36	133	4.5	380	1	US-08-120-827-51	Sequence 51, Appl
37	133	4.5	380	1	US-08-478-675-51	Sequence 51, Appl
38	132	4.5	443	2	US-08-935-450-6	Sequence 6, Appli
39	129.5	4.4	545	2	US-08-990-114-1	Sequence 1, Appli
40	129	4.4	675	4	US-08-973-273-5	Sequence 5, Appli
41	127	4.3	652	4	US-09-347-833-2	Sequence 2, Appli
42	122.5	4.1	428	4	US-09-347-833-6	Sequence 6, Appli
43	121.5	4.1	759	1	US-08-676-967-1	Sequence 1, Appli
44	121.5	4.1	759	1	US-08-676-974-1	Sequence 1, Appli
45	121.5	4.1	759	2	US-09-098-487-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-261-855-2

; Sequence 2, Application US/09261855A

; Patent No. 6255055

; GENERAL INFORMATION:

; APPLICANT: Ross, Jeffrey

; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN

; FILE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE

; FILE REFERENCE: 960296.95131

; CURRENT APPLICATION NUMBER: US/09/261,855A

; CURRENT FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 577

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-261-855-2

Query Match	74.1%	Score 2190;	DB 4;	Length 577;
Best Local Similarity	74.1%	Pred. No. 1.4e-187;		
Matches 434;	Conservative 62;	Mismatches 74;	Indels 16;	Gaps 8;
Qy	1	MNKLTYIGNLSENAAPSDLESIFPKDAKIPVSGPFLVKTGYAFVDCPDDESWALKAEIALSGK	60	
Db	1	MNKLTYIGNLSESVTPADLEKVFPAEHRKISYSGQPLVKSGYAFVDCPDHWMAMKAETPSGK	60	
Qy	61	IELHGKPIEVEHSVPKQRIKQIRNIPHLOWEVLDSLLVOYGVVSECEQVNTDSETA	120	
Db	61	VELQGRLEMEHSVPKQRSRKIQIRNIPQLRWEVLDSLLAOYGVVSECEQVNTDSETA	120	
Qy	121	VNVVTSSKDOARQALDKLNGFOLENTLKVAYIPDEMAAQNPLOQPRGR-GLGORG	179	
Db	121	VNVVTSSKDOARQALDKLNGFOLENTLKVAYIPDEMAAQNPLOQPRGR-GLGORG	179	
Qy	180	SRQSGP---GSVSKQPCDPLRLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKEN	236	
Db	177	PROGSEVAAAGAPAKQPVDPILRLVPTQYVGAIIGKEGATIRNITKQTSKIDVHRKEN	236	
Qy	237	AGAAEKSTILTPEGSTSACKSILEIMHKEAQDIKFTTEIPKILAHNNFVRLIGKEG	296	
Db	237	AGAAEKASIVHSVTEGSCACKMILEIMHKEAKDTADEVPLKILAHNNFVRLIGKEG	296	
Qy	297	RNLKTEQDTFKTITSLQELTLYNPRTITVKGNVETCAKAEEMKIRIESYENDIA	356	
Db	297	RNLKKEVDQTEFKTITSLQELTLYNPRTITVKGNVETCAKAEEMKIRIESYENDIA	356	
Qy	357	SMNLAHLIPGNNALGLFPPTSGMPPPTSGPPSANT--PPYQPEQS-ETETVHQFIP	413	
Db	357	SMNLAHLIPGNNALGLFPPTSGMPPPTSGPPSANT--PPYQPEQS-ETETVHQFIP	413	

Db	357	AMLSQSHLIGLNLAAGVLEPPASSAVPP---PPSSVTGAAPSYSSPMQAPEQMVOVFTP	413
Qy	414	ALSVGAIIGQGQHIKOLSRFASGASIKIAPAEPADAKRVMIITGPPEAQFRAQRIYCK	473
Db	414	AQVGAIIIGKGHIKOLSRFASGASIKIAPPETDSKVRMVIITGPPEAQFRAQRIYCK	473
Qy	474	IKENFVSYPKEEVKLEAHIRVPSFAAGRVIIGKGKTVNELQNLUSSAEVVVPDPTDPDEN	533
Db	474	LKENFFGPKREEVKLETHIRVPSAAGRVIIGKGKTVNELQNLTAAEYVVVPDPTDPDEN	533
Qy	534	QVVVKTGHFYACQVAQRKIQEILTVQKHQQOKALKQSFPPOSRRK	579
Db	534	QVIVKIIGHFYASOMQRKTRDLIAQVKQ-QHKQG-QSNLAQARRK	577

RESULT 2
US-08-021-608D-2
: Sequence 2, Application US/08021608D
: Patent No. 5580760
: GENERAL INFORMATION:
: APPLICANT: LEVENS, DAVID L., DUNCAN,
: APPLICANT: ROBERT C., AND AVIGAN, MARK I.
: TITLE OF INVENTION: NOVEL FUSE BINDING
: TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
:

Query Match 8.2%; Score 241; DB 1; Length 644;
Best Local Similarity 22.3%; Pred. NO. 4.2e-13;
Matches 100; Conservative 75; Mismatches 154; Indels 1

Qy	161	QONPLQ---QPRGRRLGQGRSSRQSGSVSKOKPCDPLRLLLVPTQFVGAIIIGKEGAT	21
Db	63	QKRPLEDDQDPDAKKVAPQNDSEFQTQLPPMHQOQSRSSVMTTEYKVPDGMVGFIIGRGGEQ	122
Qy	218	IRNITKQTSKIDVHRENAGAAEKSTIILSTPEGTSAAKCSLEIMHKEAQDIKFTE--	275
Db	123	ISRIQESGCKIQI-APDSGGLPERSCKLTGTPEVSQSAKRLLDQIVKGRPAPGFHHGD	181
Qy	276	---EIPKILAHNNFVRLGKKEGRNLKKIEQDQTDKITISPQLQELTLYNPRT-----I	327
Db	182	PGNAVOEIMIPASKAGLVIGKGETIKQLOERAGVKWM--IQD---GPQWTGADKPL	235
Qy	328	TVKGNVETCAKAEIEIMKKIRE-----SYENDIASMLNQAHLPGLNLAJLFPPTSGM	382
Db	236	RITGDPYKVQOAKEMVLELIRDQGGFREVRNYSR-----IGG-----NEGI	278
Qy	383	PPPTSGPPSAMTPYPPOFQSESETVHOFPALPSVGNATKQOQHILKQLSRFAGASIKIA	442
Db	279	DVP-----IPFAVGIVIGRNGEMIKKIQNDAGVRIQFK	312
Qy	443	PAEAPDAKVRMVIITGPPE-AQFKAQ-----GRIVGK-----	473
Db	313	PDDGTTPE-RIAQITGPPDRCOHAAEIIITLLRSVQAGNPGGPGGRGRCQGNWNG	371
Qy	474	----IKENFVSPKEEVKLEAHIRVPSFAAGRVIKGGKTVNBNLQNSAEVVVPRDQTP	529
Db	372	PPGGLQEFNEI-----VPTGKTGLIIGKGETIKSISQOSGARIELQRNPPP	418
Qy	530	DENDQV-VVKITGH----FYACQVAQRKI	553
Db	419	NADPNMKLFTIRGTPQOIDYAROLIEKI	447
RESULT 3			
US-08-726-160-2			
; Sequence 2, Application US/08726160			
; Patent No. 7534016			
; GENERAL INFORMATION:			
; APPLICANT: LEVENS, DAVID L., DUNCAN,			
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.			
; TITLE OF INVENTION: NOVEL FUSE BINDING			
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR			
; NUMBER OF SEQUENCES: 24			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: MORGAN & FINNEGAN			
; STREET: 345 PARK AVENUE			
; CITY: NEW YORK			
; STATE: NEW YORK			
; COUNTRY: USA			
; ZIP: 10154			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: FLOPPY DISK			
; COMPUTER: IBM PC COMPATIBLE			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: WORDPERFECT 5.1			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/726,160			
; FILING DATE: 04-OCT-1996			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 08/021,608			
; FILING DATE: 22-FEB-1993			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: WILLIAM S. FEILER			
; REGISTRATION NUMBER: 26,728			
; REFERENCE/DOCKET NUMBER: 2026-40630S1			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (212) 758-4800			
; TELEFAX: (212) 751-6849			
; INFORMATION FOR SEQ ID NO: 2:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 644			


```
QY 530 DENDQV-VVKITGH---FYACQVQVARKI 553
; : : : : : : : : : : : : : : : :
Db 419 NADNPWKLFTIRGTPQQIDYARQLIEKI 447
; : : : : : : : : : : : : : : : :

RESULT 5
US-09-261-855-22
; Sequence 22, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-261-855-22

Query Match 8.1%; Score 238; DB 4; Length 49;
Best Local Similarity 98.0%; Pred. No. 1.4e-14;
Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 281 ILAHNPVGLRGKLGKGNLKKIEQDQDTKITISPLQELTYLPERTIV 329
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 ILAHNPVGLRGKLGKGNLKKIEQDQDTKITISPLQELTYLPERTIV 49

RESULT 6
US-09-261-855-24
; Sequence 24, Application US/09261855A
; Patent No. 6255055
; GENERAL INFORMATION:
; APPLICANT: Ross, Jeffrey
; TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
; TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
; FILE REFERENCE: 960296.95131
; CURRENT APPLICATION NUMBER: US/09/261,855A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-261-855-24

Query Match 8.0%; Score 237; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 IRVPSFAAGRVIGKGGKTVNELQNLSSAEVYVPRDQTPDNDQVVKI 539
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 IRVPSFAAGRVIGKGGKTVNELQNLSSAEVYVPRDQTPDNDQVVKI 48

RESULT 7
US-08-021-608D-10
; Sequence 10, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: Amino Acid 148 (Xaa) is Met or Ile
US-08-021-608D-10

Query Match 8.0%; Score 236.5; DB 1; Length 643;
Best Local Similarity 22.3%; Pred. No. 1.1e-12;
Matches 100; Conservative 77; Mismatches 151; Indels 121; Gaps 18;

QY 161 QQNPLQ---QPRGRRGLGQRRSGSGSVSKQKPCDPLRLLLVPTQFVGAIIGEGAT 217
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 63 QKRPLEDGDQPDAAKVAPQNDSEFTQLP-PMHQQRQSVMTTEYKVPDGMVGFIIGRGEQ 121

QY 218 IRNITKQTSKIDVHRKENAGAAEKSTITLSTPEGTSAACKSILEIMHKEAQDIKTE-- 275
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 122 ISRIQESGCKIQI-APDSGGLPERSCXLTGTPESVQSAKRLLDQIVEGRGAPGFHHGD 180

QY 276 ---EIPLKILAHNFVGRITGKGRNLKKIEQDQDTKITISPLQELTYLPERT-----I 327
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 181 GPCNAVQEIIMPASKAGLVIGRGGETIKQLOERAGVQVMW--IQD---GPQNTGADKPL 234

QY 328 TVKGNVETCAKAEETIMKKIRE-----SYENDIASNNLOAHLIPGLNLAGLFPPTSGM 382
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 235 RITGDYKVKQQAEMVLELIRDQGGFREVNRNEYSR-----IGG-----NEGI 277

QY 383 PPPTSGPPSAMTPPYQFEQSETETVHQIFALPSVCAIIGKOGHILKOLSRFAGASIKIA 442
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 278 DVP-----IPFAVGIVIGRNGEMIKKIQNDAGVRIQFK 311

QY 443 PAEAPDAKVRMVIITGPE-AQFKAQ-----GRIVGK----- 473
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 312 PDDGITPE-RIAQITGPPDCQHAAEIIITDLRSVQAGNPGGPGGGRGGRGQGNWNG 370

QY 474 ----IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVYVPRDQTP 529
```

```

Db 371 PPGGLQEFNFI-----VPTGKTGLIIGGGETIKSIQSQSGARIELQRNPPP 417
QY 530 DENDQV-VVKITCH-----FYACQVAQRKI 553
Db 418 NADPNMKLFTIRGTPOQIDYARQLIEKI 446

RESULT 8
US-08-726-160-10
; Sequence 10, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726.160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION:
; OTHER INFORMATION: Amino Acid 148 (Xaa) is Met or Ile
; US-08-726-160-10

Query Match 8.0%; Score 236.5; DB 1; Length 643;
Best Local Similarity 22.3%; Pred No. 1.1e-12;
Matches 100; Conservative 77; Mismatches 151; Indels 121; Gaps 18;

QY 161 QQNPLQ---QPRRRGLGQSSRGSGSVSKQKPCDLPRLRLVPTQFVGAIIGKEGAT 217
Db 63 OKRPLEDGPQDAKKVAPQNDSEGTQLP-PMHQQORSVMTVEYKVPDGMVGFIIGRGGQ 121
QY 218 INRITKQTSKIDVIRKENAGAAEKSTITLSTPEGTSACKSILEIMHKEAQDIKFTPE-- 275
Db 122 ISRIQOESGCKIQI-APDSGGLPERSXCILGTGPSPESVQSAKRLLDQIVEKGRPAPGFHHGD 180
```

```

QY 276 ---EIPKLIAHNHNVGRILIGKEGRNLKKIEQDTRKTIISPLQELTLYNPRT-----I 327
Db 181 GPCNAVQEIWIPASKAGLVIGKGETIKQLOERAGVKVMW--IQD---GPQNTGADKPL 234
QY 328 TVKGNVETCAKAEIEIMKKIRE-----SYENDIASNMQAHLIPGLNLNGLFPPTSGM 382
Db 235 RITGDPYKVOQAKEMVLEIRDOGGFREVNEYGSR-----ICG-----NEGI 277
QY 383 PPPTSGPPSAMTPPYQPFQSETEVHQIFALSVGAIIKQGHKQLSRFAGASKIA 442
Db 278 DWP-----IPFAVGIVIGRNGEMIKKIQNDAGVRIQFK 311
QY 443 PAEAPDAKVRMVIITGPE-AQFKAO-----GRIYK-----473
Db 312 PDGTTPE-RIAQITGPPDRCCQHAELIITDLRSVQAGNPGPGGRCGRGQGNWNG 370
QY 474 ----TKEENFVSPKEVKLEAHIRVPSFAAGRVIGKGTVTNELQNLSSAEVVVPRDQTP 529
Db 371 PPGGLQEFNFI-----VPTGKTGLIIGGGETIKSIQSQSGARIELQRNPPP 417
QY 530 DENDQV-VVKITCH-----FYACQVAQRKI 553
Db 418 NADPNMKLFTIRGTPOQIDYARQLIEKI 446

RESULT 9
PCT-US94-01782-10
; Sequence 10, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
```


Db 208 D-----GFQNTGADRLRITGDPYKQAKEMVLEIRDOGGFREVNRNYSR-----IG 257
QY 367 GLNLNALGLFPPTSGMPPTSGPPSAMTPPYQPEQSETVTHQFIPALSVGAIIGKQG 426
Db 258 G-----NEGIDVP-----IPRFVGIIVGRNGE 280
QY 427 HIKLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIONDAGVRIQFKPDGTTPE-RIAQITGPPDRQCQAAEIIITDLRSVQAGNPGPG 339
QY 468 --GRIYK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTNVEL 513
Db 340 PGRGRGRGQGNWMMGPPGLQEFNF-----VPTGKTGLIIGKGETIKSI 386
QY 514 QNLSSAEVVPRDQTPDENOV-VVKITGH-----FYACQVAORKI 553
Db 387 SQSGARIELQRNPPNADPNMKLFTIRGTPOQIDYARQLIEKI 431

RESULT 12

US-08-726-160-8
; Sequence 8, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
US-08-726-160-8

Query Match 7.8%; Score 229.5; DB 1; Length 590;
Best Local Similarity 22.5%; Pred. No. 3.9e-12;

Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;
QY 202 VPTQFVGALIGKEGATIRNITKQTSKIDVHRKENAGAEKSITILSTPEGTSAAACKSIL 261
Db 91 VPDGMVFIILGGEQISRIOESCKIQI-APDSGGLPERSCLMTGTPTESVQSAKRLLD 149
QY 262 EIMHKEAQDIFKTE-----EIPKLIAHNNFVGRILIGKEGRNLKKIEQDTDTKITISPLQ 316
Db 150 QIVEKGRPAPGFHHDGDPGNAVQEIIMPASKAGLVIGKGETIKQLQERAGVKVMW--IQ 207
QY 317 ELTYLNPERT-----ITVKNVETCAKAEELMKIRE-----SYENDIASNNLOAHLIP 366
Db 208 D-----GFQNTGADRLRITGDPYKQAKEMVLEIRDOGGFREVNRNYSR-----IG 257
QY 367 GLNLNALGLFPPTSGMPPTSGPPSAMTPPYQPEQSETVTHQFIPALSVGAIIGKQG 426
Db 258 G-----NEGIDVP-----IPRFVGIIVGRNGE 280
QY 427 HIKLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIONDAGVRIQFKPDGTTPE-RIAQITGPPDRQCQAAEIIITDLRSVQAGNPGPG 339
QY 468 --GRIYK-----IKENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTNVEL 513
Db 340 PGRGRGRGQGNWMMGPPGLQEFNF-----VPTGKTGLIIGKGETIKSI 386
QY 514 QNLSSAEVVPRDQTPDENOV-VVKITGH-----FYACQVAORKI 553
Db 387 SQSGARIELQRNPPNADPNMKLFTIRGTPOQIDYARQLIEKI 431

RESULT 13

PCT-US94-01782-8
; Sequence 8, Application PC/TUS9401782
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES
; APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
; APPLICANT: HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01782
; FILING DATE: 22-FEB-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/021,608
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown

MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE: Human
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HL60
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PCT-US94-01782-8

Query Match 7.8%; Score 229.5; DB 5; Length 590;
Best Local Similarity 22.5%; Pred. No. 3.9e-12;
Matches 91; Conservative 69; Mismatches 128; Indels 117; Gaps 16;
QY 202 VPTQFYCAIIGKCATIRNITKOTQSKIDVHRKENAGAAEKSTILSTPFGTSAACKSIL 261
Db 91 VPDGMVGIIGRGEQLSRIQQESGCKIQI-APDSGGLPERSOMLTGTGPESVQSAKRLLD 149
QY 262 EIMHKEAQDIKFTPE-----EIPKILAHNNFVGRGLIGKGRNLLKTEQDITDITISPLQ 316
Db 150 QIVEKGRPAPFHHGDCPGNAQVEIMIPASKAGLVIGKGETIKQERAGVQWVM--IQ 207
QY 317 ELTYLPERT-----ITVKGNVETCAKAEIEMKKIRE-----SYENDIASMNLQAHLP 366
Db 208 D-----GPQNTGADKPLRITGDPYKVOQAKEMVLELIRDQGGFREVNRVYGSR-----IG 257
QY 367 GLNLNALGLFPPTSGMPPPTSGPPSANTPPYQFQESQETETVHQFIPALSVGAIIGKQG 426
Db 258 G-----NEGIDVP-----IPRFAVGIVIGRNGE 280
QY 427 HIKOLSRFAGASIKIAPAEAPDAKVRMVIITGPPE-AQFKAQ----- 467
Db 281 MIKKIQNDAGVRIFQKPDGDTTPE-RIAQITGPPDRCOHAAEIIITDLRSVQAGNPGGPG 339
QY 468 --GRYVGK-----IKENFVSPKEVKVLEAHIRVPSFAAGRVIGKGGKTVNEL 513
Db 340 PGGRGGRGQGNMNGPPGGLQEFNFI-----VPTGKTGLIIGKGETIKSI 386
QY 514 QNLSSAEVVVPRDQTPDENDQV-VVKITGH-----FYACOVAORKI 553
Db 387 SQSGGARIELQRPNNPPNADPNMKLFTIRGTPQOIDYARQLIEKI 431

RESULT 14
US-09-261-855-23
Sequence 23, Application US/09261855A
Patent No. 6255055
GENERAL INFORMATION:
APPLICANT: Ross, Jeffrey
TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
TITLE OF INVENTION: (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
FILE REFERENCE: 960296.95131
CURRENT APPLICATION NUMBER: US/09/261,855A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 47
TYPE: PRT
ORGANISM: Homo sapiens
US-09-261-855-23

Query Match 7.7%; Score 227; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 410 QFIPALSVGALIGKQGOHIKOLSRFAGASIKIAPAEAPDAKVRMVII 456
Db 1 QFIPALSVGALIGKQGOHIKOLSRFAGASIKIAPAEAPDAKVRMVII 47
RESULT 15
US-08-187-793-4
Sequence 4, Application US/08187793
Patent No. 5614371
GENERAL INFORMATION:
APPLICANT: POSNER, JEROME B.
APPLICANT: DARNELL, ROBERT B.
APPLICANT: FURNEAUX, HENRY M.
TITLE OF INVENTION: ANTIGEN RECOGNIZED BY ANTIBODIES ASSOCIATED
TITLE OF INVENTION: WITH PARANEOPLASTIC OPSOCLONUS AND METHODS OF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/187,793
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 691,559
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39227
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: (212) 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-187-793-4

Query Match 7.7%; Score 227; DB 1; Length 530;
Best Local Similarity 22.4%; Pred. No. 5.5e-12;
Matches 119; Conservative 77; Mismatches 179; Indels 156; Gaps 20;
QY 155 PDMAAQNPLOQPRGRRGLGQRGSSRGSGVSKQKPCDPLRLVPTQFVGAIGKE 214
Db 43 PD---SRKRPLEAP-----PEAGSTKRTNTGDDGVF-----LKVLPYAAAGSIIGK 88
QY 215 GATIRNITKOTQSKIDVHRKENA-----GAAEKSIITLSTPEGTSAACKSILMIHKEAQ 269
Db 89 GOTIVOLQKETGATIKLSKLSKDFYPGTTTERTVCLIQGTVEALNAVHGFIAEKIREMPQ 148
QY 270 DIKFTTEEIPL-----KILAHNNFVGR 290
Db 149 NVAKTEPVSILOPOTTVNPDRIKQTLPLSPSTTTTKSSPDMPTTSRANQKVIIPNSTAGL 208

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 25, 2002, 20:28:48 ; Search time 63.81 Seconds
(without alignments)
691.194 Million cell updates/sec
Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLIGNLSENAPSDLES.....VKHQOQKALQSGPPQSRRK 579
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*
1: piri:*
2: piri2:*
3: piri3:*
4: piri4:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	418.5	14.2 854 2 T23837	hypothetical prote
2	243.5	8.2 768 2 T27855	hypothetical prote
3	240	8.1 644 2 A53184	myc far upstream e
4	236.5	8.0 568 2 T49962	hypothetical prote
5	235.5	8.0 621 2 D96534	hypothetical prote
6	231.5	7.8 313 2 T48439	probable RNA-bindi
7	229.5	7.8 398 2 T41600	probable pre-mRNA
8	227	7.7 510 2 I38489	onconeural ventral
9	225.5	7.6 589 2 T19216	hypothetical prote
10	225.5	7.6 611 2 T19217	hypothetical prote
11	222	7.5 479 2 C86275	hypothetical prote
12	219.5	7.4 396 2 A41224	hnRNP protein - Af
13	213.5	7.2 680 2 T25832	hypothetical prote
14	209.5	7.1 356 2 S58529	alpha-complex prot
15	207.5	7.0 1268 2 A44125	high density lipop
16	204.5	6.9 846 2 T04533	hypothetical prote
17	204	6.9 413 2 S46109	hnRNP complex prot
18	202.5	6.9 365 2 S42471	hnRNP protein E2 -
19	202	6.8 362 2 S78515	single-stranded nu
20	202	6.8 397 2 T30168	hypothetical prote
21	199.5	6.7 1279 2 T41389	rna binding protei
22	195	6.6 649 2 E84614	probable RNA-bindi
23	194.5	6.6 1270 2 S23464	vigilin - chicken
24	190	6.4 463 2 S41495	dC stretch-binding
25	190	6.4 464 2 S43363	transformation upr
26	190	6.4 464 2 A54143	kappa-B motif-bind
27	189.5	6.4 632 2 T02627	hypothetical prote
28	184.5	6.2 1220 2 T19117	hypothetical prote
29	183	6.2 1222 2 S56030	SCP160 protein - y

ALIGNMENTS									
RESULT	1								
T23837									
hypothetical protein M88.5 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999									
C:Accession: T23837									
R:Sulston, J.									
submitted to the EMBL Data Library, June 1994									
A:Reference number: Z19806									
A:Accession: T23837									
A:Status: preliminary; translated from GB/EMBL/DBDJ									
A:Molecule type: DNA									
A:Residues: 1-854 <WILL>									
A:Cross-references: EMBL:Z34802; PIDN:CAA84338.1; GSPDB:GN00021; CESP:M88.5									
A:Experimental source: clone M88									
C:Genetics:									
A:Gene: CESP:M88.5									
A:Map position: 3									
A:Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1									
Query Match	14.2%	Score 418.5;	DB 2;	Length 854;					
Best Local Similarity	24.1%;	Pred. No. 4.5e-19;							
Matches 160;	Conservative 103;	Mismatches 215;	Indels 187;	Gaps 27;					
Qy	78	QRIRKIQIRNPPHLOWEVLDSLLVQYGVESCEQVNTDSETAVVN-----VTYSK 129							
Db	167	QOMQOQOOSNOSAHFMHQQLQAVQQO-----QAQQMHRLQAGAPINPQQQFMVPPPTMMQP 221							
Qy	130	DQARQALDKNGFOLENFTLVAYIPDEMA--AQO--NPLQQPRRRGLGQRGSSRQSGSP 185							
Db	222	QOMQQA-QQQQAQQHQMHIHQHQPMMQQAQQGYHHPHQNQHQQA-GQHQSHSQSQ 279							
Qy	186	G-----SVSKQKP-----C-----DLPLRLVPTQFVGAILGKEGATIRNITKQ 224							
Db	280	NHNQHRNHNSQSGHPHIPQNLMPRCMLKQWPIRCVEGKYHVAVLIGPNSGTIKDIASS 339							
Qy	225	TQSKID---VHRKNA--GAAEKSTILSTPEGTSAACKSILEIMHKEA--QDIKETEI 277							
Db	340	TRCRVDFVNLSSKKEKTVLGNDRILTVHGAQATKAVARILDVIQSEAVKDDVNVGADT 399							
Qy	278	PLKILAHNNFVGRLLGKGRNLKKEQDPTKTTISPQEQ-----LT-----LYNPRT 326							
Db	400	VLRRAHNQLCGRLGKAGSSIKEIMQKTGNTITVTKYIEPPGGISGLTANELLGIMERT 459							
Qy	327	ITVKG-NVETCAKABEEIMMKIRESYENDIASMNLQAHILIPGLNMLALFFPPTSGMPP- 384							
Db	460	IMVRGPSIEAVVQAEALISAKLKKCVESD-SQLRAQSMQCP-----MPPMM-MPPI 508							
Qy	385	-PTSGPPSAMTPPY-----PQFQSETETVHQ-----410							
Db	509	LPPGASSAVSAPHFIPTPVGMQVQHFASSQHLVHQNNNSFLQPGVLIQIGTTNLNQ 568							

A:Cross-references: GB:AE0051173; NID:gl1094762; PIDN:AAG29695.1; GSPDB:GN00141
C:Genetics:
A:Gene: F19C24.19
A:Map position: 1

Query Match 8.0%; Score 235.5; DB 2; Length 621;
Best Local Similarity 22.9%; Pred. No. 1.7e-07;
Matches 104; Conservative 76; Mismatches 171; Indels 103; Gaps 17;

QY 199 RLLVPTQFVGAIGKEGATRNITKQTOSKI-----D 230
||| ||| ||||| :||: ||| :
DB 22 RLCCPATRTGAIIGKGSVINHGQSVTSGKIRVDDIPVSEERVLIIAPSGKKKDES 81
||| ||| ||||| :||: ||| :
QY 231 VHRKENAG-----AEKSITITLST-----PECTSAACKSILEIMHK-----EAQDI----- 271
||| ||| ||||| :||: ||| :
DB 82 VCDSENGSEEPKQEGKSECAGTSGGDEEAPSSAQMALLRVFRIVFGDDAATVDGDEL 141
||| ||| ||||| :||: ||| :
QY 272 -KFTTEPLKILAINNFVGRILIGEGRNLLKIEODTTKTTISPLQEL--TLYNPERTIT 328
||| ||| ||||| :||: ||| :
DB 142 DKGESEGLCRMIVRGNOVDYLMKSGGRMIKIREDSGAIVRAISSTDQTPPCAFPGDVVIQ 201
||| ||| ||||| :||: ||| :
QY 329 VKGNVETCAKAEIEIMKKIRES-----YENDIASMNLQAHLPGLNLNAL 373
||| ||| ||||| :||: ||| :
DB 202 MNGKFSSVKKALLVLTNCOLESAPPTWDECFPPQGPYPPEYHSME----- 247
||| ||| ||||| :||: ||| :
QY 374 GLFPPTGMPPTSGPPSAMTPPPQPEQSESTVHQFIPALSGAIGIKOGHIKOLSR 433
||| ||| ||||| :||: ||| :
DB 248 --YHPQWDHPNPMPED--VGPPENRPVVEEVAFRLLCPADKVGSLIGKGAVVRALON 303
||| ||| ||||| :||: ||| :
QY 434 FAGASIKIAPAEAPDAKVRMVIITGPPPEAPK-----AQG---RIYGKIKKEENFVSPKEVK 487
||| ||| ||||| :||: ||| :
DB 304 ESGASIKVSD--PTHDSERIIVISARENLERRRHSLAODGVNRVHNRIVEIGF---EPSAA 359
||| ||| ||||| :||: ||| :
QY 488 LEAHRVPSFAGRVIGKGTKNELQNSAEV--VPRDPTPDNDQ--VVVKITGHFY 544
||| ||| ||||| :||: ||| :
DB 360 VVARLLVHSPYIGRLGKGHLISEMRATGASIRVFAKDQATKYESOHDEIVQVIGNL- 418
||| ||| ||||| :||: ||| :
QY 545 ACQVAQRKIOELTQVKOHOOOKAL-----QSGPP 574
||| ||| ||||| :||: ||| :
DB 419 --KTVQDALFOILCLREAMPFGRLPFGMGGPP 450
||| ||| ||||| :||: ||| :
RESULT 6
T48439
probable RNA-binding protein - Arabidopsis thaliana
N:Alternate names: protein T32M21.30
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48439
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48439
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 29/1; 108/3; 144/2; 209/1; 229/3; 261/3; 282/2
A:Note: T32M21.30

Query Match 7.8%; Score 231.5; DB 2; Length 313;
Best Local Similarity 25.1%; Pred. No. 1.1e-07;
Matches 79; Conservative 62; Mismatches 131; Indels 43; Gaps 10;

QY 183 GSPGSVSKQKP-----CDLP--LRLLVPTQFVGAIGKEGATRNITKQTOSKID 230
||| ||| ||||| :||: ||| :
DB 10 GSPBELAKSPPEHSDSEADSAEKPTRIFRVLSNAAGSVIGKGSITTEFOAKSGARIO 69
||| ||| ||||| :||: ||| :

Db 284 IGKGGFINQIRQETGATIRVNTSETDDDD--CIIFISKEFYEDQSPAVNAAIRLQQR 341

Qy 316 QELTYNPERTITVKGNVET

316 QELTYNPERTITVKGNETCAKAEETMKKIRESYENDIA

316 QELTYNPERTITVKGNVETCAKAEIEIMKKIRESYENDIASMNLQAHLPGLNALGL 375

QY 471 YGKI-KEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGTYNELONLSSAEV-VVPRDQT 528
Db 342 SEKVGDAN-----DLAISTRLLVSSQIGCLIGKGGAVISEMSRVTRNIRILOKEDV 395
QY 529 PD--ENDQVVVKITGHFVACQAKRKIQEILTQV 560
Db 396 PKIAREDEEMVQITGSPDAAMKA-----LTQV 422

RESULT 12
S41224
hRNP protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S41224
R:Stoml, H.; Matunis, M.J.; Michael, W.M.; Dreyfuss, G.
Nucleic Acids Res. 21, 1193-1198, 1993
A:Title: The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.
A:Reference number: S41224; MUID:93219080
A:Accession: S41224
A:Molecule type: mRNA
A:Residues: 1-396 <STO>
C:Superfamily: transformation upregulated nuclear protein HNRPK

Query Match 7.4%; Score 219.5; DB 2; Length 396;
Best Local Similarity 21.3%; Pred. No. 9.2e-07;
Matches 93; Conservative 76; Mismatches 166; Indels 101; Gaps 18;

QY 175 GORSSRQSGSPGVSQKPCDL--PLRLVPTQVFGAIIKKEGATINIKQTQSKIDVH 232
Db 17 GKRAEDMEDEQAFKRNRNTDMVLEKILLOSNNAGAVIGKGNKIKALRTDYNASVSV- 75
QY 233 RKNAGAAREKSITILSPETSAAKSSILEIMHKEAQDIKFTTEIPKILAHNNFVGRLI 292
Db 76 --PDSSGPERILSADIEITIGELKKIPTLEHFKGNDF--DCELRLLIHSLAGGII 131
QY 293 GKEGRNKKTEQDITKITSPLQELTYNPRTITVKNVETCAKAEIEIMKKIRES-- 350
Db 132 GVRGAKIKELREKTQT--TIKFEQCPHSTDRVLIGRPRDVRVECIKVILDLISESPV 189
QY 351 -----YEN-DIASNNL-----QAHILPGLNALGLF---PPTSGMPPTTS 387
Db 190 KGRSQYDPNFYETDYGGMFDDRRGRPH---GFSMHARGEDRMPGCGGRMPQS 246
QY 388 -----GPPSAMTPPYPOPEQSETETVHQFIPALSVGAILGQGOHIKQLSRFAGAS-- 438
Db 247 RRDYDDMSPRRGLPPPP-----GRGR-----GGSRA 274
QY 439 --TKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYKIKKEENFVSPKEEVKLEAHIRVPS 496
Db 275 RNLPLPPPPPPRGDGR-----RGRPDHYDGMGRGYRGSGFDIGGP-----VITQTVPK 326
QY 497 FAAGRVIGKGTYNELONLSSAEVVPDQTPDENDQVVVKITGHFVACQAKRKIQE- 555
Db 327 DLAGSIIGKGGRIKQIRHESGASIKI--DEPLEGSDRIITITG-----TQDQIQNA 377
QY 556 ---ILTVQKHQHOQKA 568
Db 378 RFLLNQSVKQFSEDYA 393

RESULT 13
T25832
hypothetical protein M01A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25832
R:Scheet, P.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid M01A10.
A:Reference number: Z20094

A:Accession: T25832
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <SCH>
A:Cross-references: EMBL:U88174; PIDN:AA842272.1; GSPDB:GN00019; CESP:M01A10.1
A:Experimental source: strain Bristol N2; clone M01A10
C:Genetics:
A:Gene: CESP:M01A10.1
A:Map position: 1
A:Introns: 63/2; 97/3; 121/1; 160/1; 269/3; 411/3; 512/1; 649/2

Query Match 7.2%; Score 213.5; DB 2; Length 680;
Best Local Similarity 21.3%; Pred. No. 5e-06;
Matches 73; Conservative 69; Mismatches 139; Indels 61; Gaps 9;

QY 202 VPTQFVGAIIKKEGATINIKQTQSKIDVHKNAGAAEKSIITILSPETSAAKSIL 261
Db 152 IPESVGLVIGRNGVEIOAISQKSCRVQI-VAEPSTGYRSVDIYGISENIIEVAKKLIN 210
QY 262 EIM---HKEAQD-----IKFTTEIPKILAHNNFVGRILGKGRNKKIEQ 304
Db 211 EVVARGKLSQELPCSVPOFPIPAVNSNSKVIIIPIPANKCGAIIKKGQMRKLS 270
QY 305 DTDKTIISPLQELTYNPRTITVKNVETCAKAEIEIMKKIRESYENDIASMNLQHL 364
Db 271 WTNCDFIL--IQENNIADSVKPLQITGPKVEHAHA-----LVADI 310
QY 365 IGPLNALGLFPPTSGMPPTSGPPSAMTPPYPOPEQSETETVHQFIPALSVCAIIGKQ 424
Db 311 LDGFD-----ECPPAGMAGNSPVAAMS-----LQVKVPRSTVGAINGLQ 349
QY 425 GQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQRIYKIKKEENFVSPKE 484
Db 350 GSNIKKITSNETKIQMPDDPKLMERTLVVIGNKKNKYVC-ARLQKIVEANSENANT 408
QY 485 EVKLEAHIRVPSFAAGRVIGKGTYNELONLSSAEVVPD 526
Db 409 PISL-FYMLIPASKGLVIGRGTETIRQINKSGAYCEMSRD 449

RESULT 14
S58529
alpha-complex protein 1 - human
N:Alternate names: nucleic acid-binding protein; protein PCBP-1
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C:Accession: S58529; S58523; S65678; S43489; S41378; S42472
R:Kiledjian, M.; Wang, X.; Liehaber, S.A.
EMBO J. 14, 4357-4364, 1995
A:Title: Identification of two KH domain proteins in the alpha-globin mRNA stability
A:Reference number: S58523; MUID:96016208
A:Accession: S58529
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-356 <KIL>
A:Accession: S58523
A>Status: preliminary
A:Molecule type: protein
A:Residues: 125-139; 251-265; 315-323 <K12>
R:Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A:Title: Characterisation of two major cellular poly(rC)-binding human proteins, each
A:Reference number: S65678; MUID:95331278
A:Accession: S65678
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-204, 'v', 206-356 <LEF>
A:Cross-references: EMBL:X78137; NID:g460770; PIDN:CAA55016.1; PID:g460771
A:Experimental source: AMA cells (transformed human amnion cells)
A:Note: submitted to the EMBL Data Library, March 1994
R:Aasheim, H.C.; Loukianova, T.; Deggerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994

A:Title: Tissue specific expression and cDNA structure of a human transcript encoding a
A:Reference number: S43489; MUID:94203810
A:Accession: S43489
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-298, 'H' <AAS>
A:Cross-references: EMBL:D29505; NID:9444020; PIDN:CAA82631.1; PID:9444021
C:Genetics:
A:Gene: GDB:HNRNPX
A:Cross-references: GDB:344947
C:Keywords: RNA binding

Query Match 7.1%; Score 209.5; DB 2; Length 356;
Best Local Similarity 23.4%; Pred. No. 3.5e-06;
Matches 83; Conservative 63; Mismatches 118; Indels 91; Gaps 12;

Qy 196 LPLRLVPTQVGAIGKEGATIRNITKQTSKIDVHRKENAGAEEKSITILSTPEGTSA 255

Db 14 LTRLLMHGKEVGIIGKKGESVKRIEESGARINI-----SEGNCPERIITLTGP--TNA 67

Qy 256 ACKSILEIMHKEAQDIKFT-----EEIPLKILAHNFEVGRLLIGKEGRNLKKIEODT 306

Db 68 IFKAFAMIIDKLEEDINSMTNSTAASPPVTLRLVVPATQCGSLIGKGGKIKIREST 127

Qy 307 DTKITISPLQELTYNPERTITVKGNETCAKAEIEIMKKIRES-----YE- 352

Db 128 GAQVQVA--GDMLPNSTERAITAGVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP 185

Qy 353 -----NDIASMNLOAHILPG-----LNLNALGL----- 375

Db 186 MPASSPVICAGGQDCRDAAGYPHATHDLEGPLDAYSIGQHTISPLDLAKLNOVARQQ 245

Qy 376 -----FPPTSGMPPTSGPPSMTPPYPOFEQSETEVHQF-IPALSVGAILGK 423

Db 246 SHFAMHGGTGFGIDSSPEVKG-----YNASLDASTQTHLTPNNLIGCIIGR 297

Qy 424 QGQHIKQLSRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQRIYGRKKEE 477

Db 298 QGANINEIRQMSGAQIKIANPVEGSSG--RQVITGSAASISLAQYLINARLSSE 350

RESULT 15

A44125

high density lipoprotein-binding protein, 110K - human

C:Species: Homo sapiens (man)

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999

C:Accession: A44125

R:McKnight, G.L.; Reasoner, J.; Gilbert, T.; Sundquist, K.O.; Hokland, B.; McKernan, P.A.

J. Biol. Chem. 267, 12131-12141, 1992

A:Title: Cloning and expression of a cellular high density lipoprotein-binding protein

A:Reference number: A44125; MUID:92291094

A:Accession: A44125

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1268 <MCK>

A:Cross-references: GB:M64098; GB:M83789; NID:g183891; PIDN:AAA35962.1; PID:g183892

A:Note: sequence extracted from NCBI backbone (NCBIN:106862, NCBIIP:106863)

C:Superfamily: vigilin

Query Match 7.0%; Score 207.5; DB 2; Length 1268;
Best Local Similarity 23.9%; Pred. No. 3e-05;
Matches 95; Conservative 69; Mismatches 153; Indels 81; Gaps 18;

Qy 202 VPTQFVGAIGKEGATIRNITKQTSKIDVHRKENAGAEEKSITILSTPEGTSAACKSIL 261

Db 157 IPKEHHRREVICKGKQLDLKLTATKIQIPRPPD---PSNQIKITGPKGEGARHEVL 213

Qy 262 EIMHKEAQDIKFTTEIPLKILAHNFVG---RLIGKEGRNLKKIEQDTDTKITISPLQEL 318

Db 214 LI--SAEQDKRAVERLEVEKAFHPFIAGPNYRLVG-----EIMQETGRINIPP---- 260

Qy 319 TLYNPRT-ITVKGNETCAKAEIEIMKKIRESYENDIASMNL-----QAHLIPGLNLNA 372

Db 261 --PSVNRTEIVTGEKEQLAQAVARI-KKIYEKKKKTTTIAVEVKKSQHKKVIGPKGNS 317

Qy 373 LGLFPPTSG---MPPPTS-----GPP---SMTTPPYQFEQSETETV-----HQ 410

Db 318 LQEIILERTGVSVVEIIPSDSISSETVILRGPEKLQALTEVYAKANSFTVSSVAAPSWLHR 377

Qy 411 FIPALSVGAIIGKQGHKQLSRFAGASIKIAPAPADAKVRMVIITGPPEAQFKAQGR I 470

Db 378 F-----IIGKKQNLAKITQ-QMPKVHIEFTEGED-----KITLEGPTEDVNVAAEQEI 424

Qy 471 YGKIKE-----ENFVSPKEVEVKLEAHIRVPSFAAGRVIGKGTVNELQNLSSAEVVVPRD 526

Db 425 EGMVADLINRMDYVEINIDHKFHRHL-----IGKSGANINRIKQYKVSVRIP-- 472

Qy 527 QTPDENQVVVKITGHFYACQVAQKIQEILLTQVKOHQ 564

Db 473 --PDSEKSNLIRIEGDPQGVQQAQKRELLLELASRMENER 508

Search completed: January 25, 2002, 22:17:32

Job time: 6524 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 22:12:48 ; Search time 52.07 Seconds
(without alignments)
407.700 Million cell updates/sec

Title: US-09-685-696-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQKALQSGPPQSRKK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	227	7.7	7.7	510	1	NOA1_HUMAN	P51513 homo sapien
2	224.5	7.6	7.6	339	1	PCB3_HUMAN	P57721 homo sapien
3	219.5	7.4	7.4	339	1	PCB3_MOUSE	P57722 mus musculus
4	213.5	7.2	7.2	403	1	PCB4_HUMAN	P57723 homo sapien
5	212.5	7.2	7.2	403	1	PCB4_MOUSE	P57724 mus musculus
6	210.5	7.1	7.1	606	1	Y475_ARATH	P58223 arabidopsis
7	209.5	7.1	7.1	356	1	PCB1_HUMAN	P51365 homo sapien
8	209.5	7.1	7.1	356	1	PCB1_RABIT	P19048 oryctolagus
9	207.5	7.0	7.0	1268	1	VGLN_HUMAN	O00341 homo sapien
10	204	6.9	6.9	413	1	PBR2_YEAST	P38151 saccharomyc
11	202.5	6.9	6.9	365	1	PCB2_HUMAN	Q15366 homo sapien
12	202	6.8	6.8	362	1	PCB2_MOUSE	Q61990 mus musculus
13	194.5	6.6	6.6	1270	1	VGLN_CHICK	P81021 gallus gall
14	190	6.4	6.4	463	1	ROK_HUMAN	Q07244 homo sapien
15	190	6.4	6.4	463	1	ROK_RABIT	P19049 oryctolagus
16	190	6.4	6.4	464	1	ROK_MOUSE	Q60577 mus musculus
17	183	6.2	6.2	1222	1	S160_YEAST	P08105 saccharomyc
18	170.5	5.8	6.29	629	1	PAB2_ARATH	P42731 arabidopsis
19	164.5	5.6	6.44	644	1	PAB4_HUMAN	Q13310 homo sapien
20	160.5	5.4	6.36	636	1	PAB1_MOUSE	P29341 mus musculus
21	159.5	5.4	4.70	1	NR54_HUMAN	Q15233 homo sapien	
22	159.5	5.4	6.36	636	1	PAB1_HUMAN	P11940 homo sapien
23	156	5.3	6.33	633	1	PABP_XENLA	P20965 xenopus lae
24	152.5	5.2	3.81	1	YBP2_YEAST	P38199 saccharomyc	
25	150	5.1	3.44	1	SFR6_HUMAN	Q13247 homo sapien	
26	147.5	5.0	7.47	1	FCA_ARATH	O04425 arabidopsis	
27	145.5	4.9	4.14	1	NSR1_YEAST	P27476 saccharomyc	
28	145	4.9	4.14	1	NSP3_YEAST	Q01560 saccharomyc	
29	139.5	4.7	5.22	1	PAB2_HUMAN	Q15097 homo sapien	
30	138.5	4.7	4.24	1	S3B4_HUMAN	Q15427 homo sapien	
31	135.5	4.6	3.59	1	ELV2_HUMAN	Q12926 homo sapien	
32	135.5	4.6	3.60	1	ELV2_MOUSE	Q60899 mus musculus	
33	135	4.6	3.59	1	ELV3_HUMAN	Q14576 homo sapien	

34	135	4.6	367	1	ELV3_MOUSE	Q06900 mus musculu
35	135	4.6	494	1	SFR4_HUMAN	Q08170 homo sapien
36	134	4.5	285	1	CABA_MOUSE	Q09020 mus musculu
37	134	4.5	307	1	SXL_MEGSC	O01671 megaselia s
38	133	4.5	373	1	ELV4_RAT	O09032 rattus norv
39	133	4.5	380	1	ELV4_HUMAN	P26378 homo sapien
40	133	4.5	385	1	ELV4_MOUSE	Q061701 mus musculu
41	130	4.4	633	1	ROR_HUMAN	O43390 homo sapien
42	129.5	4.4	6359	1	BACC_BACLI	O68008 b bacitraci
43	129	4.4	653	1	PABP_SCHPO	P31209 schizosacch
44	128.5	4.3	660	1	PAB3_ARATH	O64380 arabidopsis
45	128	4.3	576	1	PABP_YEAST	P04147 saccharomyc

ALIGNMENTS

RESULT 1	NOA1_HUMAN	STANDARD;	PRT;	510 AA.
ID	NOA1_HUMAN			
AC	P51513;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	ONCONEURAL VENTRAL ANTIGEN-1 (NOVA-1) (PARANEOPLASTIC RI ANTIGEN)			
DE	(VENTRAL NEURON-SPECIFIC PROTEIN 1).			
GN	NOVA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cerebellum, and Hippocampus;			
RX	MEDLINE=94000830; PubMed=8398153;			
RA	Buckanovich R.J., Posner J.B., Darnell R.B.;			
RT	"Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding			
RT	protein and is specifically expressed in the developing motor			
RT	system."			
RL	Neuron 11:657-672(1993).			
RN	[2]			
RP	SEQUENCE OF 1-34 FROM N.A.			
RC	TISSUE=Fetal brain;			
RA	Dmitrenko V.V., Garifulin O.M., Shostak K.A., Smikodub A.I.,			
RA	Kavsan V.M.;			
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC			
CC	SUBSET OF DEVELOPING NEURONS.			
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-!- TISSUE SPECIFICITY: BRAIN.			
CC	-!- DISEASE: ANTIBODIES ARE SEEN AGAINST THIS ANTIGEN IN THE PATIENTS			
CC	SUFFERING WITH PARANEOPLASTIC OPSOCLONUS-ATAXIA (POA). POA IS A			
CC	DISORDER IN WHICH ABNORMAL MOTOR CONTROL OF THE EYES, TRUNK AND			
CC	LIMBS DEVELOP IN WOMEN WITH BREAST OR SMALL LUNG CANCER. POA			
CC	PATIENTS SUFFER FROM OPSOCLONUS, A CHAOTIC EYE MOVEMENT DISORDER			
CC	ATTRIBUTABLE TO A LACK OF INHIBITION OF BURST NEURONS IN THE BRAIN			
CC	STEM: MYOCLONUS, A DISORDER PERHAPS ATTRIBUTING TO A FAILURE OF			
CC	INHIBITORY CONTROL OVER SPINAL MOTOR NEURONS; AND TRUNCAL ATAXIA,			
CC	REFERABLE TO CEREBELLAR DYSFUNCTION; THERE ARE NO CORTICAL OR			
CC	SENSORY DEFICITS.			
CC	-!- SIMILARITY: CONTAINS 3 KH DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U04840; AAA16022.1;			
DR	EMBL; Z70771; CAA94810.1;			
DR	MIM; 602157; ..			


```

RX MEDLINE=20396135; PubMed=10936052;
RA Makeyev A.V.; Liebhauer S.A.;
RT "Identification of two novel mammalian genes establishes a subfamily
RL of KH-domain RNA-binding proteins.";
RL Genomics 67:301-316(2000)
CC -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF176327; AAG09238.1; -
DR MGD: MGI:1890470; Pcbp3.
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PS00084; KH_TYPE_1; 3.
DR Nucleic acid protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW Repeat.
FT DOMAIN 13 63 KH 1.
FT DOMAIN 97 150 KH 2.
FT DOMAIN 261 313 KH 3.
SQ SEQUENCE 339 AA; 35958 MW; A6F1C7C176A64F9C CRC64;

Query Match 7.4%; Score 219.5; DB 1; Length 339;
Best Local Similarity 20.6%; Pred. No. 1e-06;
Matches 75; Conservative 52; Mismatches 108; Indels 129; Gaps 7;

QY 277 IPLKILAHNNFVGRIGREGRNLKKIEQDQDTKTIISPLQELTYNPTVKGNVETC 336
D 14 LTIRLLMHGKEVGSIIKKGETVKKMRESGARINISEGN-----CPEIVITIGTDAI 68
QY 337 AKAEIEIMKKIESYENDIASNMNLAHLIPLGNLNLALGLFPPTSGMPPPTSGPPSAMTPP 396
D 69 KFAFIAMAYKEEDILINSN-----SPATSKPP----- 97
QY 397 YPQFEQSETEVHQFIPALSVGAIIGKOGHIKQLSRFAGASIKIAPAEAPDAKVRMVI 456
D 98 -----VTLRVVPASQCSGLIGKGGSKIKERESTGAQVQVAGOMLPNSTERAVTI 148
QY 457 TGPPPAQFK-----AQGRIV----- 471
D 149 SCTPDALFQCVKQICVVMLESPPKGTIPYRKPACTPVIFAGGQAYTIQOYAIHPDQ 208
QY 472 -----GKIEENFVSPKEVKLEA-----HIRV 494
D 209 LTKHLQAMQOTFPFPLGOTNPAFPEKELPLHSSEEAQNLMQSSGLDASPPASTHELT 268
QY 495 PSFAAGRVIGKGGKYNELQNLSSREVVPVPRDTPDENDQVVKVTGHEFYACQVQRKIQ 554
D 269 PNDLIGCLIGROGTKINEIROMSGAQIKIANTEGSSERQIT--ITGTPTANISLAOYLIN 326
QY 555 EILT 558
D 327 ARLT 330

RESULT 4
PCB4_HUMAN
ID PCB4_HUMAN STANDARD; PRT; 403 AA.
AC P57723.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Created)

```

```

DE POLY(RC)-BINDING PROTEIN 4 (ALPHA-CP4).
GN PCBP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20396135; PubMed=10936052;
RA Makeyev A.V.; Liebhauer S.A.;
RT "Identification of two novel mammalian genes establishes a subfamily
RL of KH-domain RNA-binding proteins.";
RL Genomics 67:301-316(2000)
CC -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF176330; AAG09241.1; -
DR InterPro: IPR000958; KH.
DR Pfam: PF00013; KH-domain; 3.
DR SMART: SM00322; KH; 3.
DR PROSITE: PS00084; KH_TYPE_1; 3.
DR Nucleic acid protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW Repeat.
FT DOMAIN 17 67 KH 1.
FT DOMAIN 101 154 KH 2.
FT DOMAIN 241 293 KH 3.
SQ SEQUENCE 403 AA; 41481 MW; 3D99F762A9471265 CRC64;

Query Match 7.2%; Score 213.5; DB 1; Length 403;
Best Local Similarity 25.6%; Pred. No. 3.1e-06;
Matches 81; Conservative 61; Mismatches 126; Indels 49; Gaps 13;

QY 183 GSPGSVSKQKPCD--LPLRLVPTQFVGAIIGKEGATIRNITKQTSKIDVHRKENAGAA 240
D 3 GSDGGLLEEPESLITLTLRLMHGKEVGSIIKKGETVKKIRIQSSARITI-----SEGSC 58
QY 241 EKSITILSTPETSACKSILEIMHKEAQDI-----KFTTEIPLKILAHNNFVGRIL 291
D 59 PERITIT--GSTAAVFHAVSMIAFKLDELCAAPANGNVSRPPVTLRLVIPASQCSGL 116
QY 292 IGKEGRNLKKIEQDQDTKTIISPLQELTYNPTVKGNVETCAKABEIEIMKKIESY 351
D 117 IGKAGTKIKIEIRETTGAQVQA--GDLLPNSTERAVTVSGVPDAIILCVRQICAVILESP 174
QY 352 ENDIASNMNLAHLIPLGNLNLALGLFPPTSGMPPPTSGPPSAMTP-----PY 397
D 175 PK---GATPYH--PSLSLGTV-LLSANQGF--SVQGYGAVTPAEVTKLQQLSSHAVPF 226
QY 398 -----PQFEQ--SETETVHQFIPALSVGAIIGKOGHIKQLSRFAGASIKIAPAEAPDAK 450
D 227 ATPSVVPGLDPGTQTSSQEFVLPNDLIGCVIGRQSGKISIRIOMSGAHKIKIG-NAQEGAG 285
QY 451 VRMVIITGPPPEAQFKAQ 467
D 286 ERHVTITGSPVSIALAQ 302

RESULT 5
PCB4_MOUSE
ID PCB4_MOUSE STANDARD; PRT; 403 AA.
AC P57724;
DT 20-AUG-2001 (Rel. 40, Created)

```


Db	68	IFKAFAMIIOKLEEDINSSMTNSTAASRPVTVLRLVVPATQCSGLIGKGCKIKEIREST	127
Qy	307	DTKITISPLQELTYLNPERITVYKGNVETCAKAEIEIMKKIRES-----YE-	352
Db	128	GAQVOVA--GDMLPNSTERAITIATVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP	185
Qy	353	-----NDIASMNLQAHLPG-----LNLNALGL-----	375
Db	186	MPASSVICAGGQRCSDAVGYPHATHDLEGGPDLDAYISQGOHTISPLDLAKLNQVAAQO	245
Qy	376	-----FPPTSGMPPPTSGPPSAMTPPYPQPEQSESETVTHQF-IPALSVGAIGK	423
Db	246	SHEAMMHGGTGAGIDSSSEVKG-----YVASILDASTQTTHELTIPNNLIGCIIGR	297
Qy	424	QGQHIKQLSFAGASIKIA-PACAPADAKVRWVLIITGPPEAQFAQGRIVYKKKEE	477
Db	298	QGANINEIRMQSAQKIAMPVEGSSG--RQVTTIGTSAASISLAQYLIANRLSS	350

RESULT 8

```

PCB01_RABIT
ID      PCB1_RABIT      STANDARD;      PRT;      356 AA.
AC      O19048;
DT      20-AUG-2001 (Rel. 40, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      POLY(RC)-BINDING PROTEIN 1 (ALPHA-CPI) (HNRNP-E1).
GN      PCBPI.
OS      Oryctolagus cuniculus (Rabbit).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NC      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE-99207122; Pubmed-10101180;
RT      Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT      "Tissue-specific transcriptional regulation of alternative rabbit
RT      15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RT      Nucleic Acids Res. 27:1828-1836(1999).
CC      -1- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT BINDS
CC      PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC      -1- SIMILARITY: CONTAINS 3 KH DOMAINS.
-----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).

```

RESULT 9

ID VGLN_HUMAN STANDARD; PRT: 1268 AA.
 AC Q09C41; Q9UCY3;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VIGILIN (HIGH DENSITY LIPOPROTEIN-BINDING PROTEIN) (HDL-BINDING
 DE PROTEIN).
 DE HDLBP OR HBP OR VGL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92291094; PubMed=1318310;
 RA McKnight G.L., Reasoner J., Gilbert T., Sundquist K.O., Hokland B.,
 RA McKernan P.A., Champeone J., Johnson C.J., Bailey M.C., Holly R.,
 RA O'Hara P.J., Oram J.F.;
 RT "Cloning and expression of a cellular high density lipoprotein-binding
 RT protein that is up-regulated by cholesterol loading of cells.";
 RT J. Biol. Chem. 267:12131-12141(1992).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=96184515; PubMed=8605996;
 RA Kugler S., Grunweller A., Probst C., Klingner M., Muller P.K.,
 RA Kruse C.;
 RT "Viglin contains a functional nuclear localisation sequence and is
 RT present in both the cytoplasm and the nucleus.";
 RT FEBS Lett. 382:330-334(1996).
 CC -1- FUNCTION: APPEARS TO PLAY A ROLE IN CELL STEROL METABOLISM.
 CC IT MAY FUNCTION TO PROTECT CELLS FROM OVER-ACCUMULATION OF
 CC CHOLESTEROL.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
 CC -1- SIMILARITY: CONTAINS 14 KH DOMAINS.
 CC -1- SIMILARITY: STRONG, TO C.ELEGANS C08H9.2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collabora-
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no
 CC modified and this statement is not removed, usage by and for commer-
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64098; AAA35962.1; -
 CC MIM: 142695; -
 CC InterPro: IPR000958; KH.
 CC Pfam: PF00013; KH-domain; 14.
 CC SMART: SM00322; KH; 14.
 CC PROSITE: PS50084; KH_TYPE_1; 14.
 CC HDL: Lipid transport; Cholesterol metabolism; RNA-binding; Repeat:

Query Match 6.8%; Score 202; DB 1; Length 362;

Best Local Similarity 24.2%; Pred. No. 1.3e-05;

Matches 87; Conservative 60; Mismatches 117; Indels 96; Gaps 14;

QY 196 LPLRLVPTQFVGAIGEGATIRNITKQTSKIDVHKKENAGAAEKSTITLSTPEGTS 255

DB 14 LTRILLMHGKVGSIIGKSGESVKMRBESGARINI-----SEGNGPERIITLAGP--TNA 67

QY 256 ACKSILETMHKEAQDIKFT-----EEIPLKLIAHNNFVGRILGKGRNLKIEQDT 306

DB 68 IPKAFAMIIDKLEEDISSMTWNSTASRPVTLRLVPASQCSGLIGKGGCKIKIREST 127

QY 307 DTKITISPLQELTYLNPRTIVKG---NVETCAKAEIEIMKKI----- 347

DB 128 GAQVQVA--GDMLPNSTERAITIAGIQSIIECVKQICVMLESPPKGVTIYRKPSS 185

QY 348 -----RESYENDIASMNLQAHILPGLNLNALGLFPPTSGMP----- 383

DB 186 PVIFAGQDRYSTGSDASF---PHTTSMCLN-----PDLEGPPLEAVTIQGYAIPOP 237

QY 384 -----PPTSGPP--SAMTPPYPO-----PEQSETVYHQF--IPALSVG 418

DB 238 DITKLHLAQMOQSHFPMTHGNTGFGSIESSPEVGYWAGLDASAQTTSHELTIPNDLIG 297

QY 419 ALIIGKOGHIKOLSFAGASIKIA--PAAPDAKVRWVITGPPEAQFKAQGRIVYKIEE 477

DB 298 CLIQGAKINEIROMSGAQIKIANPVEG--STDROVITGSAISLSLAQYILNVLRSSE 355

RESULT 13

VLGN_CHICK

ID VGLN_CHICK STANDARD; PRT; 1270 AA.

AC P81021;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE VIGILIN.

GN HDLBP OR VGL.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

ON NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryonic sternal cartilage;

RX MEDLINE=92298988; PubMed=1606952;

RA Schmidt C., Henkel B., Poeschl E., Zorbas H., Purschke W.G.,

RA Gloc T.R., Mueller P.K.;

RT "Complete cDNA sequence of chicken vigilin, a novel protein with

amplified and evolutionary conserved domains.";

RL Eur. J. Biochem. 206:625-634(1992).

CC -!- SUBCELLULAR LOCATION: INTRACELLULAR.

CC -!- SIMILARITY: CONTAINS 14 KH DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; X65292; CAA46387.1; -

DR InterPro; IPR000958; KH.

DR Pfam; PF00013; KH-domain; 14.

DR SMART; SM00322; KH; 14.

DR PROSITE; PS50084; KH_TYPE_1; 14.

KW RNA-binding; Repeat.

FT DOMAIN 150 188 KH 1.

FT DOMAIN 219 260 KH 2.

FT DOMAIN 291 333 KH 3.

FT DOMAIN 360 402 KH 4.

FT DOMAIN 431 473 KH 5.

FT DOMAIN 504 545 KH 6.

FT DOMAIN 577 619 KH 7.

FT DOMAIN 651 693 KH 8.

FT DOMAIN 724 766 KH 9.

FT DOMAIN 798 840 KH 10.

FT DOMAIN 872 913 KH 11.

FT DOMAIN 970 1012 KH 12.

FT DOMAIN 1051 1093 KH 13.

FT DOMAIN 1126 1168 KH 14.

SQ SEQUENCE 1270 AA; 142220 MW; 0A8844F91F9B8619 CRC64;

Query Match

Best Local Similarity 21.8%; Pred. No. 0.00023;

Matches 97; Conservative 85; Mismatches 168; Indels 95; Gaps 17;

QY 182 QGSGSV--SKQPCDPLR-----LLVPTQFVGAIGEGATIRNITKQTSKI-- 229

DB 483 EGDPOGVQQAQKELLELSASMENERTKOLJIEQKPHRTIIGKGERIREIREKFEVIIN 542

QY 230 ---DVHKKENAGAAEKSTITLSTPEGTSACKSILIMHKEAQDIKFTTEIPLKILAHNN 286

DB 543 FPDPAHKSD-----IVQLRGPKNVEKCTKYMOKMVADLVENSFSISVPFIKQFHN 594

QY 287 FVGRILGKGRNLKIEQDTDTKITISPLQELTYLNPRTITVKGNVETCAKAEIEIMKK 346

DB 595 ----IIGKGANIKKIREESNTKIDLPGREQAT----QOLLSQGREQTVKLLRHRL-- 644

QY 347 IRESYENDIASMNLQAHILPGLNLNAL---GLFP-----PTSG----- 381

DB 645 ---AQKELANITEVEVSIPLKLSLIGAKGRFIRSIIMECGGVHIFPTEGSSATVT 701

QY 382 ---MPPPTSGPPSAMTPPYQFQSOSETETV-----HOFIPALSVAIGIGKQGHKQL 431

DB 702 IRAQPTWRKPRSSCCTWAEKQTSYTVDLRAKPYHKF-----LIGKGGGIRKV 753

QY 432 SRFAGASIKIAPAEADPAKVRWVITGPPEAQFKAQGRIVYKIE--ENFVSPKEEVKLEA 490

DB 754 RNTGARIIFPTSEDKQDE--LITIMGTEAEVKAQKEALELIKLNLDNVVEDSMVDPKH 811

QY 491 HIRVPSFAAGRVIGKGTVNELQNLSAEVYVPRDQTPDENQVYVVKITGHFYACQVAQ 550

DB 812 H-----RHFVIRGQVLRLEIADEYGVVMVRLPTVSGTQSDKVTLK--GAKDCVEAAK 861

QY 551 RKIQEILTQVKHQKQKALQSGPPQ 575

DB 862 KRQIEIIDL---EAQVTIECTIPQ 883

RESULT 14

ROK_HUMAN

ID ROK_HUMAN STANDARD; PRT; 463 AA.

AC Q07244; Q15671;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K (HNRNP K) (DC-STRETCH

DE BINDING PROTEIN) (CSBP) (TRANSFORMATION UPREGULATED NUCLEAR PROTEIN)

DE (TUNP).

GN HNRPK OR HNRNPK.

OS Homo sapiens (Human), and Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

ON NCBI_TaxID=9606, 10116;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Human.

RX MEDLINE=92107165; PubMed=1729596;

RA Matunis M.J., Michael W.M., Dreyfuss G.;

RT "Characterization and primary structure of the poly(C)-binding

RT heterogeneous nuclear ribonucleoprotein complex K protein.";
RL Mol. Cell. Biol. 12:164-171(1992).
RN [2]
RP SPECIES=Human.
RC SEQUENCE FROM N.A.
RX MEDLINE=94149726; PubMed=8107114;
RA Dejgaard K., Jeffers H., Rasmussen H.H., Madsen P., Kruse T.A.,
RA Gesser B., Nielsen H., Celis J.E.;
RT "Identification, molecular cloning, expression and chromosome mapping
RT of a family of transformation upregulated hnRNP-K proteins derived by
RT alternative splicing.";
RL J. Mol. Biol. 236:33-48(1994).
RN [3]
RN SEQUENCE FROM N.A.
RP SPECIES=Rat; STRAIN=DONRYU; TISSUE=Liver;
RX MEDLINE=94173662; PubMed=8127654;
RA Ito K., Sato K., Endo H.;
RT "Cloning and characterization of a single-stranded DNA binding
RT protein that specifically recognizes deoxycytidine stretch.";
RL Nucleic Acids Res. 22:53-58(1994).
CC -1- FUNCTION: ONE OF THE MAJOR PRE-MRNA-BINDING PROTEINS. BINDS
CC TENACIOUSLY TO POLY(C) SEQUENCES. LIKELY TO PLAY A ROLE IN
CC THE NUCLEAR METABOLISM OF HNRNAS, PARTICULARLY FOR PRE-MRNAS THAT
CC CONTAIN CYTIDINE-RICH SEQUENCES. CAN ALSO BIND POLY(C) SINGLE-
CC STRANDED DNA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOPLASM.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S74678; AAB20770.1; -
DR EMBL; X72727; CAA51267.1; -
DR EMBL; D17711; BAA04566.1; -
DR PIR; A42058; A42058.
DR PIR; S41495; S41495.
DR MIM; 600712; -
DR InterPro; IPR000958; KH.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS50084; KH_TYPE_1; 3.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat; DNA-binding;
KW Phosphorylation; Alternative splicing.
FT DOMAIN 42 92
FT KH 1.
FT RNA-BINDING (RGG-BOX).
FT DOMAIN 144 197
FT KH 2.
FT DOMAIN 236 273
FT KH 3.
FT DOMAIN 387 439
FT KH 4.
FT DOMAIN 54 421
FT KH 5.
FT REPEAT 399 421
FT 1-1.
FT REPEAT 245 329
FT 1-2.
FT DOMAIN 245 329
FT 2 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 245 329
FT 2-1.
FT DOMAIN 59 407
FT 5 X 4 AA REPEATS OF G-X-G-G.
FT REPEAT 59 62
FT 3-1.
FT REPEAT 257 260
FT 3-2.
FT REPEAT 267 270
FT 3-3.
FT REPEAT 295 298
FT 3-4.
FT REPEAT 404 407
FT 3-5.
FT DOMAIN 289 294
FT POLY-PRO.
FT DOMAIN 310 315
FT POLY-PRO.
FT VARSPLIC 459 463
FT SKKFF -> ADVEGF (IN ISOFORM 2).
FT CONFLICT 32 32
FT A -> D (IN REF. 2).
SQ SEQUENCE 463 AA; 50976 MW; 0F70EE169B2A064A CRC64;

Query Match 6.4%; Score 190; DB 1; Length 463;
Best Local Similarity 20.7%; Pred. No. 0.0001;
Matches 87; Conservative 54; Mismatches 131; Indels 148; Gaps 13;
QY 198 LRLVPTQFVGAIGREGATIRNITKQTSKIDVHRKENAGAAEKSTITLSTPGTSAAC 257
DB 45 LRLLOSKNAGAVIGKGNKIKALRTDYNASVS---PDSGPERILSISADIETGEIL 101
QY 258 KSILETMHKEAQ--DIKFTTEIP-----LKIHAHNFVGRIG 293
DB 102 KKIPILEGLQLPSPATSQLPLESDAVCLNYQHYKGSDFCELRLLHQSLAGGIIG 161
QY 294 KEGRNKKIEQDITDTKITISPLQELTLNPERTITVKNVETCAKAEIEIMKKIRES--- 350
DB 162 VKGAKIKELRENTQT--TIKLFQECCHPHSTDRVVLIGGKPDVVVECIKIILDLISESPIK 219
QY 351 -----YENDIASMNLQAHILPGLNALGLF---PPTSG---MPP----- 384
DB 220 GRAQYDPNFYDITYGGFTMMFDDRRGRPVGFPMRGGGDFRMPGGRGPRMPPSRRD 279
QY 385 -----PTSGGPPS-----AMTPPVQ-----F 400
DB 280 YDMSPRGPPPPPPGRRGGRSARNLPLPPPPPPRGGDL MAYDRGRGDRYDGMVGF 339
QY 401 EQSET-----ETVHOFIPALSV 417
DB 340 SADETWDSAITWSPSEWOMAYEPQGGSGYDYSVAGRGSGYDLGGPIITQTVPKDLA 399
QY 418 GAIIGKQGHKQLSRFAGASIKI-APAEAPDAKVRVVIITGPPEAQFKAQGRYIGKIKE 476
DB 400 GSIIIGKGGQIKOIRHESGASIKIDEPLG--SEDRITITITGTQDIQNAQYLLQNSVKQ 457
RESULT 15
ROK_RABIT
ID ROK_RABIT STANDARD; PRT; 463 AA.
AC O19049;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K (HNRNP K).
GN HNRPK.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99207122; PubMed=10101190;
RA Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT "Tissue-specific translational regulation of alternative rabbit
RT 15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL Nucleic Acids Res. 27:1828-1836(1999).
CC -1- FUNCTION: ONE OF THE MAJOR PRE-MRNA-BINDING PROTEINS. BINDS
CC TENACIOUSLY TO POLY(C) SEQUENCES. LIKELY TO PLAY A ROLE IN THE
CC NUCLEAR METABOLISM OF HNRNAS, PARTICULARLY FOR PRE-MRNAS THAT
CC CONTAIN CYTIDINE-RICH SEQUENCES. CAN ALSO BIND POLY(C) SINGLE-
CC STRANDED DNA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOPLASM (BY
CC SIMILARITY).
CC -1- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ003024; CAA05815.1; -

```
DR InterPro; IPR000958; KH.  
DR Pfam; PF00013; KH-domain; 3.  
DR SMART; SM00322; KH; 3.  
DR PROSITE; PS00084; KH TYPE 1; 3.  
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat; DNA-binding;  
KW Phosphorylation.  
FT DOMAIN 42 92 KH 1.  
FT DOMAIN 144 197 KH 2.  
FT DOMAIN 236 273 RNA-BINDING (RGG-BOX).  
FT DOMAIN 387 439 KH 3.  
FT DOMAIN 54 421 2 X 22 AA APPROXIMATE REPEATS.  
FT REPEAT 54 76 1-1.  
FT REPEAT 399 421 1-2.  
FT DOMAIN 245 329 2 X 6 AA REPEATS OF D-R-R-G-R-P.  
FT REPEAT 245 250 2-1.  
FT REPEAT 324 329 2-2.  
FT DOMAIN 59 407 5 X 4 AA REPEATS OF G-X-G-G.  
FT REPEAT 59 62 3-1.  
FT REPEAT 257 260 3-2.  
FT REPEAT 267 270 3-3.  
FT REPEAT 295 298 3-4.  
FT REPEAT 404 407 3-5.  
FT DOMAIN 289 294 POLY-PRO.  
FT DOMAIN 308 315 POLY-PRO.  
SQ SEQUENCE 463 AA; 50960 MW; A62A92E645EFE1B6 CRC64;  
  
Query Match 6.4%; Score 190; DB 1; Length 463;  
Best Local Similarity 20.7%; Pred. No. 0.0001;  
Matches 87; Conservative 54; Mismatches 131; Indels 148; Gaps 13;  
  
Qy 198 LRLVPTQFVGAIGKEGATIRNITKQSKIDVHRKENAGAAEKSIITLSTPGTSAAC 257  
||: : : ||: ||: | : : : : : : : : : : : : : : : : : : : : : : : :  
Db 45 LRLIQSKNAGAVICKGCKNIKALRTDYNASVSV---PDSSGPERILSISADIETIGEL 101  
  
Qy 258 KSILEIMHKEAQ--DIKTEEIP-----LKTIAHNNFVGRIG 293  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 102 KKIIPTEEGQLPSPTATSQLPLESDAVECLNYQHFKGSDFDCELRLLIHQSLAGGIIG 161  
  
Qy 294 KEGRLKKIEODTKITISPLQELTYNPRTITVGNVETCAKAEEMKKIRES--- 350  
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
Db 162 VGAKIKELRENTQT--TIKLFQECCHSTDRVVLIGKPRVVECIKIILDLISEPIK 219  
  
Qy 351 -----YENDIASMNLQAHLPGLNLNALGLF---PPTSG---MPP----- 384  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 220 GRAQPYDPNFYDGYGFTMMFDRGRVPVGFPMRGGRGDFRMPGGRPPPSRRD 279  
  
Qy 385 -----PTSGPPS-----AMTPPYDQ-----F 400  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 280 YDDMSFRGRPPPPPPRGGRGGRARNLPLPPPPPPRGDLMAYDRGRPGDRYDGMVGF 339  
  
Qy 401 EQSET-----ETVHQFIPALSV 417  
|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 340 SADETWSAIDTWSSEWQAYEPQGGSGYDYSYAGGRSGYDGLGGLPITTTQVTIPKDLA 399  
  
Qy 418 GAIIKGQGHKQLSRFAGASIKI-APAEAPDAKVRMWIITGPPEAQKAGQRIYVKIKE 476  
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
Db 400 GSIICKGQRIKQIRHESGASIKIDEPLG--SEDRITITGTQDQIQNAQYLLQNSVKQ 457
```

Search completed: January 25, 2002, 22:20:22
Job time: 454 sec

